

AL/CF-TR-1995-0142

UNITED STATES AIR FORCE  
ARMSTRONG LABORATORY



THE INFLUENCE OF STROBOSCOPIC  
AUDITORY STIMULI ON VISUAL APPARENT  
MOTION PERCEPTION (U)

Michael W. Haas

CREW SYSTEMS DIRECTORATE  
HUMAN ENGINEERING DIVISION  
WRIGHT-PATTERSON AFB, OH 45433-7022

NOVEMBER 1995

DTIC QUALITY INSPECTED 2

FINAL REPORT FOR THE PERIOD OCTOBER 1989 TO NOVEMBER 1995

19970730 079

Approved for public release; distribution is unlimited

Crew Systems Directorate  
Human Engineering Division  
2255 H Street  
Wright-Patterson AFB, OH 45433-7022

## NOTICES

When US Government drawings, specifications, or other data are used for any purpose other than a definitely related Government procurement operation, the Government thereby incurs no responsibility nor any obligation whatsoever, and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication or otherwise, as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

Please do not request copies of this report from the Armstrong Laboratory. Additional copies may be purchased from:

National Technical Information Service  
5285 Port Royal Road  
Springfield, Virginia 22161

Federal Government agencies and their contractors registered with the Defense Technical Information Center should direct requests for copies of this report to:

Defense Technical Information Center  
8725 John J. Kingman Road, Suite 0944  
Ft. Belvoir, Virginia 22060-6218

## DISCLAIMER

This Technical Report is published as received and has not been edited by the Technical Editing Staff of the Armstrong Laboratory.

## TECHNICAL REVIEW AND APPROVAL

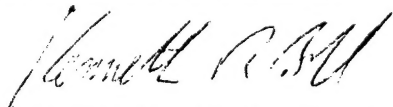
AL/CF-TR-1995-0142

This report has been reviewed by the Office of Public Affairs (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS, it will be available to the general public, including foreign nations.

The voluntary informed consent of the subjects used in this research was obtained as required by Air Force Instruction 40-402.

This technical report has been reviewed and is approved for publication.

## FOR THE COMMANDER



**KENNETH R. BOFF**, Chief  
Human Engineering Division  
Armstrong Laboratory

# REPORT DOCUMENTATION PAGE

Form Approved  
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)

2. REPORT DATE  
Nov 1995

3. REPORT TYPE AND DATES COVERED  
Final, Oct 89-Nov 95

4. TITLE AND SUBTITLE

The Influence of Stroboscopic Auditory Stimuli  
on Visual Apparent Motion Perception (U)

5. FUNDING NUMBERS

PE 62202F  
PR ILIR  
TA HE  
WU 01

6. AUTHOR(S)

Michael W. Haas

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

8. PERFORMING ORGANIZATION  
REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Armstrong Laboratory, Crew Systems Directorate  
Human Engineering Division  
Human Systems Center  
Air Force Materiel Command  
Wright-Patterson AFB OH 45433-7022

10. SPONSORING/MONITORING  
AGENCY REPORT NUMBER  
AL/CF-TR-1995-0142

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

The influence of moving auditory stimuli on visual apparent motion perception was empirically investigated. The experiments investigated perceptual organization of visual stimuli driven by inter-stimulus interval (ISI) and angular extent in the presence of moving and non-moving auditory stimuli. Characteristics of, and refinements to, a new model of visual-auditory apparent motion perception are described. Dynamic characteristics of an auditory localizer, which was used in a large portion of the work, were evaluated to assess the ability to generalize the experimental results to other conditions.

Influence characteristics of contemporaneous moving auditory stimuli on angular-extent-driven and ISI-driven visual perceptions were found to be cognitive in nature, small in magnitude, susceptible to perceptual hysteresis, and existed only when the visual-based perception was ambiguous. The small auditory influence over visual apparent motion perception was found to affect performance of a complex task, that task being tracking of an intermittent visual-auditory target, relative to tracking of an intermittent visual-only target. The auditory influence was affected by characteristics of the target movement and caused a reduction in the power spectral density of correlated and non-correlated tracking error between 0.1Hz and 0.5Hz.

14. SUBJECT TERMS

visual perception, auditory perception, intersensory  
perception, manual operation, control attention

15. NUMBER OF PAGES

248

16. PRICE CODE

17. SECURITY CLASSIFICATION  
OF REPORT

UNCLASSIFIED

18. SECURITY CLASSIFICATION  
OF THIS PAGE

UNCLASSIFIED

19. SECURITY CLASSIFICATION  
OF ABSTRACT

UNCLASSIFIED

20. LIMITATION OF ABSTRACT

UNLIMITED

THIS PAGE INTENTIONALLY LEFT BLANK



## **ABSTRACT**

### **THE INFLUENCE OF STROBOSCOPIC AUDITORY STIMULI ON VISUAL APPARENT MOTION PERCEPTION**

by Michael William Haas

The understanding of human motion perception has fundamental importance. This report increases that understanding by investigating the influence of moving auditory stimuli on visual apparent motion perception. Within this report, a new model of visual-auditory apparent motion perception is described based upon literature regarding *intra*-sensory and *inter*-sensory apparent motion perception. Characteristics of, and refinements to, this new model are founded upon experiments described in this report. These experiments investigate perceptual organization of visual stimuli driven by inter-stimulus interval (ISI) and angular extent in the presence of moving and non-moving auditory stimuli.

Influence characteristics of contemporaneous moving auditory stimuli on angular-extent-driven and ISI-driven visual perceptions were measured in five experiments. The influence was found to be cognitive in nature, small in magnitude, susceptible to perceptual hysteresis, and existed only when the visual-based perception was ambiguous. When angular-extent-driven stroboscopic visual stimuli of approximately 5° horizontal extent were augmented with moving auditory stimuli, an increase was measured in the angular extent capable of sustaining stable perceptions. Within ISIs ranging from 83ms to 150ms, an increase was measured in the ISI capable of sustaining stable perceptions of ISI-driven stroboscopic visual stimuli when augmented with moving auditory stimuli.

The small auditory influence over visual apparent motion perception was found to affect performance of a complex task, that task being tracking of an intermittent visual-auditory target, relative to tracking of an intermittent visual-only target. The auditory influence was affected by characteristics of the target movement and caused a reduction in the power spectral density of correlated and non-correlated tracking error between 0.1Hz and 0.5Hz.

Dynamic characteristics of an auditory localizer, which was used in a large portion of the work, were evaluated to assess the ability to generalize the experimental results in this report to other conditions. The auditory localizer generated auditory stimuli that elicited velocity discriminations of 14% of velocities between  $20^{\circ}\text{sec}^{-1}$  and  $100^{\circ}\text{sec}^{-1}$ . The minimum auditory movement angle measured using the auditory localizer was  $8.1^{\circ}$  at  $90^{\circ}\text{sec}^{-1}$ , which differed from previous studies in the literature using real stimuli by less than 3%.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	General introduction . . . . .	1
1.2	Research focus: Motion perception . . . . .	2
1.2.1	Research focus hierarchy . . . . .	2
1.2.2	Types of motion perception . . . . .	2
1.3	Research applications: From tactical aircraft cockpits to the totally virtual environment . . . . .	4
1.3.1	Virtual environment technology . . . . .	6
1.3.2	Virtual environment research needs . . . . .	6
<b>2</b>	<b>Objectives of research</b>	<b>7</b>
2.1	Inter-sensory apparent motion perception . . . . .	7
2.2	Overview model of inter-sensory motion perception . . . . .	8
2.3	Overview of the research . . . . .	9
2.4	Organization of the report . . . . .	10
<b>3</b>	<b>Review of the literature</b>	<b>13</b>
3.1	Intra-modal literature . . . . .	13
3.1.1	Korte's law-based models of visual and auditory apparent mo- tion perception . . . . .	13
3.1.2	Physiological structures . . . . .	20
3.1.3	Mathematical/computational models of visual motion detection and perception . . . . .	25
3.1.4	Psychophysical factors affecting human motion perception . . .	32
3.1.5	<i>Intra</i> -modal and <i>inter</i> -modal models of visual and auditory ap- parent motion perception . . . . .	50
3.2	Inter-modal literature . . . . .	57
3.2.1	Physiological basis for visual and auditory interactions . . . .	57
3.2.2	Visual and auditory perceptual interactions . . . . .	58
3.2.3	A process model of auditory influence over visual motion per- ception . . . . .	62
3.3	Summary of the literature review . . . . .	65
<b>4</b>	<b>Apparatus and methods description</b>	<b>67</b>
4.1	Speaker/LED stroboscopic subject station . . . . .	67
4.2	Localizer/CRT subject station . . . . .	70

4.2.1	Hardware components and characteristics . . . . .	70
4.2.2	Visual and auditory display synchrony . . . . .	73
4.2.3	Physical layout and controls . . . . .	74
<b>5</b>	<b>Spatial correspondence influences</b>	<b>76</b>
5.1	Experiment one: Effects of moving auditory stimuli on horizontal and vertical correspondence thresholds within a visual apparent motion display . . . . .	76
5.1.1	Aims . . . . .	79
5.1.2	Independent variables . . . . .	81
5.1.3	Dependent variable . . . . .	81
5.1.4	Other conditions . . . . .	81
5.1.5	Subjects . . . . .	82
5.1.6	Task . . . . .	82
5.1.7	Results and discussion . . . . .	83
<b>6</b>	<b>Localizer dynamic evaluation</b>	<b>93</b>
6.1	Localizability of static stimuli . . . . .	93
6.2	Experiment two: Auditory velocity discrimination . . . . .	97
6.2.1	Aims . . . . .	97
6.2.2	Independent variables . . . . .	98
6.2.3	Other conditions . . . . .	98
6.2.4	Subjects . . . . .	98
6.2.5	Task . . . . .	98
6.2.6	Dependent variable . . . . .	99
6.2.7	Results and discussion . . . . .	99
6.3	Experiment three: Minimum audible movement angle . . . . .	110
6.3.1	Aims . . . . .	110
6.3.2	Independent variables . . . . .	110
6.3.3	Dependent variable . . . . .	111
6.3.4	Other Conditions . . . . .	111
6.3.5	Subjects . . . . .	111
6.3.6	Task . . . . .	111
6.3.7	Results and discussion . . . . .	112
6.4	Summary of the auditory localizer validation investigation . . . . .	114
<b>7</b>	<b>Spatially-linked influences</b>	<b>115</b>
7.1	Experiment four: Effects of moving virtual auditory stimuli on horizontal and vertical correspondence thresholds within a visual apparent motion display . . . . .	115
7.1.1	Aims . . . . .	116
7.1.2	Independent variables . . . . .	117
7.1.3	Dependent variable . . . . .	118
7.1.4	Other conditions . . . . .	118
7.1.5	Subjects . . . . .	118
7.1.6	Task . . . . .	118

7.1.7	Results and discussion . . . . .	119
7.1.8	Contrasts with experiment one . . . . .	131
7.2	Experiment five: An effect of auditory motion on correspondence in a visual apparent motion display, a second investigation . . . . .	135
7.2.1	Aims . . . . .	136
7.2.2	Independent variables . . . . .	137
7.2.3	Dependent variable . . . . .	138
7.2.4	Other conditions . . . . .	138
7.2.5	Subjects . . . . .	138
7.2.6	Task . . . . .	138
7.2.7	Results and discussion . . . . .	139
<b>8</b>	<b>Temporal correspondence influences</b>	<b>144</b>
8.1	Experiment six: An effect of auditory motion on a temporally multi-stable visual display . . . . .	144
8.1.1	Aims . . . . .	146
8.1.2	Independent variables . . . . .	148
8.1.3	Dependent variable . . . . .	149
8.1.4	Other conditions . . . . .	149
8.1.5	Subjects . . . . .	151
8.1.6	Task . . . . .	151
8.1.7	Results and discussion . . . . .	152
8.2	Experiment seven: Auditory influence on a visually multi-stable display, a second investigation . . . . .	161
8.2.1	Aims . . . . .	161
8.2.2	Independent variables . . . . .	162
8.2.3	Dependent variable . . . . .	162
8.2.4	Other conditions . . . . .	162
8.2.5	Subjects . . . . .	163
8.2.6	Task . . . . .	163
8.2.7	Results and discussion . . . . .	164
8.3	Contrasts between temporal correspondence influence experimental results . . . . .	170
<b>9</b>	<b>Manual tracking of visual-auditory targets</b>	<b>172</b>
9.1	Experiment eight: Manual tracking of an intermittent auditory and visual target . . . . .	172
9.1.1	Aims . . . . .	172
9.1.2	Task . . . . .	176
9.1.3	Independent variables . . . . .	182
9.1.4	Dependent variable . . . . .	182
9.1.5	Other conditions . . . . .	182
9.1.6	Subjects . . . . .	184
9.1.7	Results and discussion . . . . .	184
<b>10</b>	<b>Conclusions, discussions, and recommendations</b>	<b>203</b>

10.1	Conclusions . . . . .	203
10.1.1	Moving auditory stimuli can influence angular-extent-driven visual perceptual organization . . . . .	203
10.1.2	Moving auditory stimuli can influence ISI-driven visual perceptual organization . . . . .	204
10.1.3	Temporal factors can mask inter-sensory influence effects on perceptual organization . . . . .	204
10.1.4	Pursuit tracking of a visual target can be affected by auditory target augmentation . . . . .	206
10.2	Summary of conclusions . . . . .	207
10.3	Implications of conclusions on the inter-modal process model . . . . .	208
10.4	Recommendations . . . . .	211
10.4.1	Visual and auditory linkage characteristics . . . . .	211
10.4.2	Aiding wide off-boresight tracking with auditory displays . . . . .	211
10.4.3	Tracking of non-linked visual-auditory targets . . . . .	212
10.4.4	Interaction between intermittent, multi-sensory target movement and tracking performance . . . . .	212
<b>A</b>	<b>Sample subject consent form</b>	<b>224</b>
<b>B</b>	<b>Instructions to subjects</b>	<b>227</b>

# List of Figures

1.1	Structured description of research topic area . . . . .	3
1.2	Generic auditory and visual virtual environment generation . . . . .	5
2.1	Simple inter-sensory motion perception overview model . . . . .	8
2.2	Overview model highlighting research focus . . . . .	9
3.1	Basic structural model of visual and auditory motion perception . . . . .	16
3.2	Graph of linear representation of Korte's second law . . . . .	18
3.3	Graph of multiplicative representation Korte's second law . . . . .	19
3.4	Visual or auditory perceptual strength as a function of $I$ . . . . .	21
3.5	Visual or auditory perceptual strength as a function of $E$ . . . . .	21
3.6	Motion pathways within the visual system . . . . .	23
3.7	Physiological auditory pathways . . . . .	24
3.8	Computational model based on neuronal activity . . . . .	26
3.9	Elaborated Reichardt detector model . . . . .	27
3.10	Adelson and Bergen detector model . . . . .	28
3.11	Watson and Ahumada detector model . . . . .	29
3.12	Shunting inhibition detector model . . . . .	29
3.13	Continuous visual motion report model . . . . .	36
3.14	Broken visual motion report model . . . . .	37
3.15	Continuous visual motion report model in two variables . . . . .	38
3.16	Broken visual motion report model in two variables . . . . .	39
3.17	Continuous auditory motion report model . . . . .	43
3.18	Broken auditory motion report model . . . . .	44
3.19	Continuous auditory motion report model in two variables . . . . .	45
3.20	Broken auditory motion report model in two variables . . . . .	46
3.21	Summary model kernel . . . . .	52
3.22	Spatially-combined Reichardt kernels . . . . .	54
3.23	Visual and auditory apparent motion perception process model . . . . .	56
3.24	Components that may have contributed to auditory influence over vi- sual motion perception . . . . .	63
3.25	An inter-modal process model . . . . .	64
4.1	Speaker/LED Stroboscopic Apparatus . . . . .	68
4.2	Speaker/LED Architecture . . . . .	69
4.3	Localizer/CRT Subject Station . . . . .	71
4.4	Auditory localizer internal structure . . . . .	72

5.1	One dot-two dot visual display . . . . .	77
5.2	Potential perceptual organizations . . . . .	78
5.3	Process model combining horizontal and vertical apparent motion . . . . .	80
5.4	Mean perceptual threshold extent . . . . .	84
5.5	Mean perceptual threshold extent, t-block . . . . .	86
5.6	Mean perceptual threshold extent; ascending trials, t-block . . . . .	87
5.7	Mean perceptual threshold extent: Descending trials, t-block . . . . .	88
5.8	Threshold extent for each subject with moving auditory stimuli . . . . .	89
5.9	Threshold extent for each subject with stationary auditory stimuli . . . . .	90
5.10	$E_H$ for each subject with stationary and moving auditory stimuli . . . . .	91
6.1	Auditory localizer power spectral densities . . . . .	96
6.2	Auditory virtual environment used in experiment two. . . . .	97
6.3	Velocity discrimination plotted against horizontal extent and delta velocity . . . . .	99
6.4	Smoothed mesh depiction of velocity discrimination plotted against horizontal extent and delta velocity . . . . .	100
6.5	Velocity discrimination plotted against horizontal extent . . . . .	101
6.6	Velocity discrimination plotted against base velocity . . . . .	102
6.7	Velocity discrimination plotted against delta velocity . . . . .	103
6.8	Base velocity and delta velocity affected velocity discrimination . . . . .	107
6.9	Smoothed mesh plot of velocity discrimination against base velocity and delta velocity . . . . .	108
7.1	Mean horizontal extent thresholds at $E_V = 2^\circ$ . . . . .	120
7.2	Mean thresholds at $E_V = 5^\circ$ . . . . .	121
7.3	Temporal spread of thresholds for $E_V = 5^\circ$ . . . . .	124
7.4	Temporal spread of thresholds from only ascending trials for $E_V = 5^\circ$ . . . . .	125
7.5	Temporal spread of thresholds from only descending trials for $E_V = 5^\circ$ . . . . .	126
7.6	Temporal spread of thresholds for $E_V = 2^\circ$ . . . . .	128
7.7	Temporal spread of thresholds from only ascending trials for $E_V = 2^\circ$ . . . . .	129
7.8	Temporal spread of thresholds from only descending trials for $E_V = 2^\circ$ . . . . .	130
7.9	Cumulative distribution of thresholds obtained with $S_{mode} = S$ . . . . .	133
7.10	Cumulative distribution of thresholds obtained with $S_{mode} = M$ . . . . .	134
7.11	Mean percentage of vertical motion reports . . . . .	140
7.12	Time blocked mean percentage of vertical motion reports . . . . .	142
8.1	Multi-stable two-frame visual display . . . . .	145
8.2	Mechanism model supporting perception of the 3-dot/3-dot visual stimulus . . . . .	146
8.3	Process model supporting perception of the 3-dot/3-dot visual stimulus . . . . .	148
8.4	Auditory presentation configurations in experiment six . . . . .	150
8.5	Threshold ISI at $3^\circ$ horizontal extent . . . . .	153
8.6	Threshold ISI at $9^\circ$ horizontal extent . . . . .	154
8.7	Re-categorized threshold ISI at $3^\circ$ horizontal extent . . . . .	157
8.8	Re-categorized threshold ISI at $9^\circ$ horizontal extent . . . . .	158

8.9	Percentage of <i>group</i> motion reports . . . . .	165
8.10	Blocked <i>group</i> motion reports with moving auditory presentations . . .	167
8.11	Blocked <i>group</i> motion reports with static auditory presentations . . . .	168
9.1	Block diagram of the elements involved in the pursuit tracking task . .	174
9.2	Target time history . . . . .	178
9.3	Target movement power spectral density . . . . .	179
9.4	Target power spectral density of lower frequencies . . . . .	180
9.5	Target movement power spectral density with latched intermittent periods	181
9.6	Visual target tracking symbology . . . . .	183
9.7	Mean RMS tracking error . . . . .	185
9.8	Tracking errors during intermittent periods . . . . .	186
9.9	Tracking error 400ms following the intermittent periods . . . . .	187
9.10	Cursor power spectral density from static auditory trials . . . . .	189
9.11	Cursor power spectral density from moving auditory trials . . . . .	190
9.12	Parsimonious single input/multiple output model of the pursuit track- ing task . . . . .	191
9.13	Correlated error within individual moving and static auditory trials . .	192
9.14	Non-correlated error within individual moving and static auditory trials	193
9.15	Averaged correlated error from target movement with intermittent periods	194
9.16	Averaged non-correlated error from target movement with intermittent periods . . . . .	195
9.17	Relationship between velocity difference and inter-modal influence . . .	200
10.1	Modified inter-modal process model . . . . .	210



# List of Tables

2.1	Experimental descriptions . . . . .	10
3.1	Visual perception categories . . . . .	33
3.2	Auditory perception categories . . . . .	41
5.1	Summary of analyses of variance for $E_H$ . . . . .	84
5.2	Summary of analyses of variance for $E_{H-asc}$ . . . . .	85
5.3	Summary of analyses of variance for $E_{H-dsc}$ . . . . .	85
6.1	Summary of localizer static performance . . . . .	94
6.2	Velocity discrimination analyses of variance . . . . .	104
6.3	Comparison of empirical MAMAs . . . . .	114
7.1	$E_H$ analysis of variance . . . . .	122
7.2	$E_{H-asc}$ analysis of variance . . . . .	122
7.3	$E_{H-dsc}$ analysis of variance . . . . .	122
7.4	Average thresholds obtained in experiment one and experiment four . .	132
7.5	Analysis of variance under 4 values of $E_H$ . . . . .	140
7.6	Analysis of variance under 3 values of $E_H$ . . . . .	141
7.7	$E_H$ analyses of variance . . . . .	143
8.1	Analyses of variance results, ascending trials . . . . .	152
8.2	Analyses of variance results, descending trials . . . . .	155
8.3	Average ISI analyses of variance results . . . . .	155
8.4	Re-categorized analysis of variance summary from ascending trials . . .	156
8.5	Re-categorized analysis of variance summary from descending trials . .	159
8.6	Re-categorized analysis of variance summary from average of ascending and descending trials . . . . .	159
8.7	Analysis of variance - percentage of group motion reports . . . . .	166
8.8	Effect of temporal block on <i>group</i> motion reports . . . . .	170
9.1	Intermittent period onset timing, position, and velocity characteristics .	177
9.2	Tracking error analysis of variance results using 20-second time-slices .	185
9.3	Intermittent period tracking error, analysis of variance summary . . . .	187
9.4	Post 400ms tracking error, analysis of variance summary . . . . .	188
9.5	Effect auditory presentation mode on error power spectral density . . .	196
9.6	Regression analysis independent variable description . . . . .	197
9.7	Tracking error difference linear regression analysis . . . . .	198

9.8	Tracking error difference linear regression analysis . . . . .	199
9.9	Tracking error difference linear regression analysis . . . . .	199
9.10	Secondary tracking error difference linear regression analysis . . . . .	200

## ACKNOWLEDGMENTS

This report was submitted as a dissertation to the Faculty of Engineering and Applied Science, Institute of Sound and Vibration Research, University of Southampton, United Kingdom, as partial fulfillment for the degree of Doctor of Philosophy. The research described in this report was initiated as a Laboratory Director's Fund project approved by Col. Charles P. Hatsell, former Commander of the Harry G. Armstrong Aerospace Medical Research Laboratory.

Researchers from both the Armstrong Laboratory and the University of Southampton supported this research in various ways. I would like to thank Dr. Kenneth R. Boff, Chief of the Human Engineering Division, Armstrong Laboratory, for opening the door onto the possibility of this research program by introducing me to the eventual chairperson of my supervisory committee at the University of Southampton, Professor M. J. Griffin. I am also indebted to Dr. Boff for his continual support and encouragement throughout the final years of this research program.

I am grateful to Mr. Charles Bates, prior Director of the Human Engineering Division, Armstrong Laboratory, for his support during the initial years of this research program.

I would like to thank Dr. Thomas Furness III, my prior supervisor and Chief of the Visual Displays Branch, Armstrong Laboratory, for providing encouragement when I needed it and for leading by example.

I would also like to acknowledge my supervisory committee, composed of Professor M. J. Griffin, Dr. I. H. Flindell, and Professor P. A. Nelson, for their critical reviews, insightful questions, and constructive comments. I would like to thank my external examiner, Professor R. Meddis, Head of the Human Sciences Department, University of Loughborough, United Kingdom, for his thorough review of the dissertation research.

# Chapter 1

## Introduction

### 1.1 General introduction

The challenges involved in interfacing humans with control and display devices are increasing due to advancements in control and display device capability as well as the growth of application complexity. Foley described one example of this challenge as being the interface between human and advanced computational devices [31]. Foley stated that the interface between humans and advanced computational devices may be limiting the productivity and efficiency obtainable through automation and machine intelligence [31].

Multi-sensory interface concepts may provide enhancements in the human-device interface by capitalizing on the human's innate ability to integrate, assimilate, and fuse multiple sensory experiences simultaneously. Multi-sensory interface development, guided by known characteristics of the human perceptual system, may result in the development of more powerful interface concepts.

The engineering development of multi-sensory interface concepts may be guided by a fundamental understanding of how the human receives and processes multi-sensory information. It would be an oversimplification to expect that all processing of multi-sensory information within the human occurs in separate sensory channels with no influence, or interaction, between the sensory channels. Thus a fundamental understanding of both *intra*-sensory and *inter*-sensory human characteristics may be required.

The empirical work described within this report uniquely adds to the fundamental understanding of the influence of auditory perception on visual perception and thus, to the body of overall knowledge concerning *inter*-sensory perception.

## 1.2 Research focus: Motion perception

The ability to perceive motion may be necessary for many animals. Not only do humans perceive motion but by actively interacting with their environment, also create motion within their environment [104]. Motion can be perceived by humans in many sense modalities, including vision and audition. It is reasonable to state that motion is an essential part of the human perceptual environment.

A complete understanding of the perception of motion in humans has been, and continues to be, elusive, even though research has been conducted in this area by many scientists beginning in 1875 by Exner. Fundamental research of human motion perception continues to be performed as evidenced by the number of papers and books published on the subject every year.

The research described within this report focuses on the sensory/perceptual processing characteristics of human motion perception elicited by the presentation of contemporaneous, stroboscopic, visual and auditory stimuli.

### 1.2.1 Research focus hierarchy

Understanding the sensory integration characteristics of motion perception elicited from contemporaneous visual and auditory stimuli is the dominant theme of this research. The research theme can be captured visually as several levels in a hierarchy. The research theme hierarchy, at multiple levels of focus, is shown in Figure 1.1. At the most general level, this research can be described as human-machine interface investigation. While that title completely encompasses the subject matter of this research, it does not describe the purpose, goals, or approaches taken in the research. Many different levels of description can be focused, each of which differs in completeness and specificity. A more encompassing specification of the research purpose can be obtained by traversing the hierarchy of descriptions shown in Figure 1.1.

### 1.2.2 Types of motion perception

Research topics comprising the area of motion perception can be organized in many ways. One such organization was utilized in the proceedings of a NATO symposium on motion perception [138]. The symposium consisted of the following topics; eye movement and motion perception, visual space perception through motion, thresholds of motion perception, psychophysics of motion perception, visual localization and eye movements, linear self motion perception, neural substrates of the visual perception of movement, and implications of recent developments in dynamic spatial orientation and visual resolution for vehicle guidance. Other, more general, organizations for visual motion perception can be constructed utilizing classifications of the perceptions themselves.

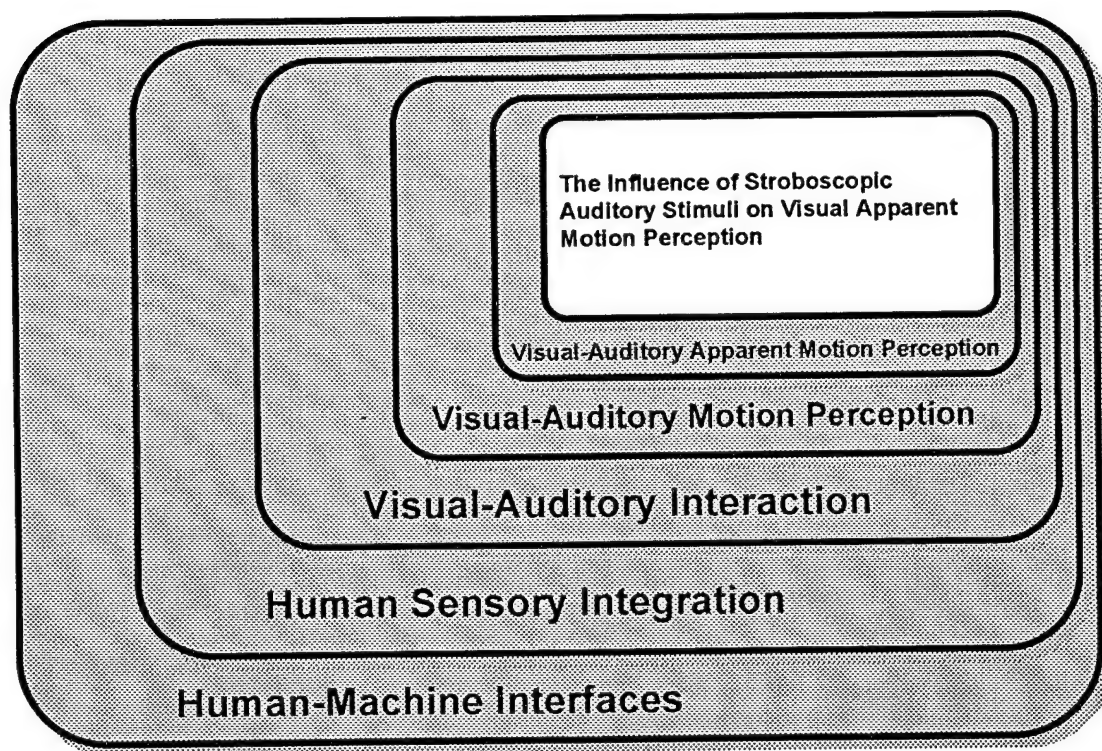


Figure 1.1: The theme of this research could be captured visually as a hierarchy of topic descriptions. The topic areas shown on this figure range from general to specific.

Sekuler presented an organization of motion perception consisting of biological motion, self motion, and apparent motion as perceptual classifications [104]. This type of organization was stimulus independent. Goldstein described several methods to stimulate visual motion perceptions that led to an organization that was stimulus dependent [36]. Hochberg utilized the same stimulus dependent organization [63]. The organization used by Goldstein and Hochberg is described in more detail below.

**Real movement:** Real movement is defined as the perception of motion that is elicited from movement within the physical environment.

**Autokinetic movement:** Autokinetic motion perception occurs in humans when points of light are viewed against a dark background. An example of this type of motion perception can be elicited by observing stars against a dark night sky. The observer of this type of motion perception may report that the points of light are moving when the points of light are physically stationary.

**Induced movement:** Induced movement is the perception of motion of a stationary stimulus that may be elicited when a moving visual stimulus is presented in conjunction with a stationary stimulus. A common example of this is the perception of movement of the full moon when it is viewed through wispy clouds moving in the night sky.

**Motion aftereffects:** Motion aftereffects are perceptions of motion created by previously occurring moving stimuli.

**Stroboscopic movement:** The discrete change of stimuli from frame to frame is, by definition, stroboscopic stimuli. Apparent motion is defined as the perception of motion that may be elicited from stroboscopic stimuli.

Five distinct types of visual apparent motion effects were described within the literature [13]. The five types were Alpha, Beta, Phi, Gamma, and Delta [13]. Alpha was the label for the perceived change in the size of stroboscopic stimuli. Beta was the label for perceived smooth and continuous movement. Phi was the label for perceived *objectless* movement. Gamma was the label of perceived size change under luminance changes. Delta was the label for perceived reverse movement caused by luminance changes.

### 1.3 Research applications: From tactical aircraft cockpits to the totally virtual environment

The application of multi-sensory interface concepts may be best accomplished using a combination of non-virtual (or conventional) and virtual control and display devices. The experience derived from the use of a combination of virtual and non-virtual devices can be termed virtual reality. The perceptual space created by

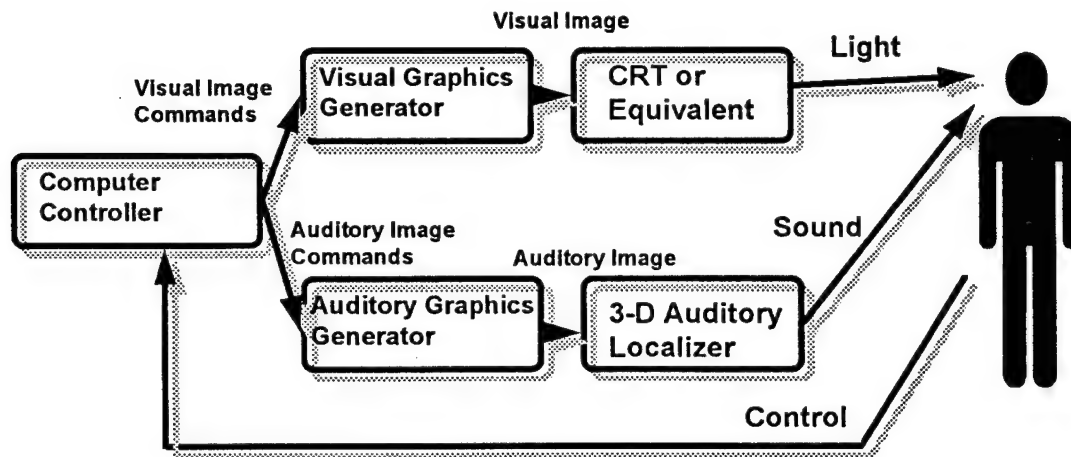


Figure 1.2: The virtual environment utilizes computer controlled generation of electronically-formatted video and audio signals which are transformed into visual and auditory stimuli.

this experience had been termed artificial reality, the virtual environment, the synthetic environment, and cyberspace [71] [26].

The virtual environment concept takes advantage of the fact that humans experience reality through a combination of sensory stimulation and retrieval of internally stored representations of the environment. The experience of a current environment is formed by the continual processing of energy emanating from the environment that is transformed through the sense organs and transmitted within the human through the central nervous system. The processing of this energy is complex and adaptive.

Virtual environments can create artificial realities for humans. In essence, to the human in the virtual environment, one reality exists outside the virtual environment and a second reality exists within the virtual environment. A generic block diagram of a visual and auditory virtual environment generator is shown in Figure 1.2. The quality of the virtual reality created by the virtual environment generator is dependent on the fidelity of the sensory stimulation. If the intent is to emulate a naturally occurring environment, the virtual environment must accurately recreate the sensory stimulation of the naturally occurring environment and allow the human to naturally interact with, and affect, the virtual environment as if it were real.



### 1.3.1 Virtual environment technology

The technology to emulate naturally occurring environments within a virtual environment does not presently exist. However, devices enabling the creation of limited virtual experiences do exist. The application of virtual environment generation for scientific visualization and design support has been identified by several industrial corporations, such as Autodesk, Sense8, Vermont Microsystems, and Dataquest [55]. A specific application example is research conducted by Doll concerning auditory localizers as directional cueing devices [22].

In addition, devices which can generate and portray virtual images, such as helmet-mounted displays, helmet-mounted head, hand, and eye line-of-sight trackers, three-dimensional auditory displays, and tactile stimulation devices, have been developed and evaluated by academic, industrial, and military institutions [33], [46], [31], [131], [137], [135], [136], [109].

Perceptual research impacting the design, development and application of virtual environment technology in the areas of vision, audition, and proprioception has been performed for many years, and is continuing, within the disciplines of psychology, human factors, and industrial engineering.

### 1.3.2 Virtual environment research needs

An area of research that demands significant attention at the present time is the perceptual integration aspects of displaying virtual and non-virtual information. While technology components for virtual displays have been developed, and are typically stroboscopic in nature, a complete understanding of the integration mechanisms evoked in the human by multi-sensory stimuli does not currently exist. This need was documented by Furness in 1986 when he described several human factors issues to be addressed in the design and development of the *Super Cockpit* [32]. In Furness' paper, he posed questions such as How should portrayal modalities be used together? [32]. The research described within this report addresses a portion of that question, specifically, how might contemporaneous stroboscopic auditory stimuli affect visual apparent motion perception.

This report adds to the fundamental understanding of how vision and audition may interact within the human and thus, provides additional insight into the development of virtual environment technology. This insight, in turn, may advance the application of virtual environment technology into systems such as Furness' *Super Cockpit*.

## Chapter 2

# Objectives of research

The objectives of this research are formed around two issues. The first issue is that presently, scientific knowledge of inter-sensory apparent motion perception is limited. The second issue is that the use of visual and auditory stroboscopic stimuli within totally-virtual and virtually-augmented interfaces is increasing, and may be unavoidable, due to the use of digital computers for image generation, digital signal processing, and raster-based methods of image portrayal.

### 2.1 Inter-sensory apparent motion perception

The objectives of this research are threefold.

**Objective 1:** The first objective of this research is to evaluate the potential existence of moving auditory stimuli influence on visual apparent motion perception.

**Objective 2:** The second objective of this research is to illuminate and quantify several characteristics of the strength of this influence.

**Objective 3:** The third objective of this research is to begin to determine if these characteristics transform into performance effects within complex task environments.

The perspective taken within this research is one of viewing the visual-auditory perceptual interaction as a product of an integrated sensory-perceptual process. The perspective taken in this research enables the interactions and influences of the auditory perceptual system on visual perception to be illuminated and investigated. This perspective can be captured in a simple model. That model is described in the following sections.

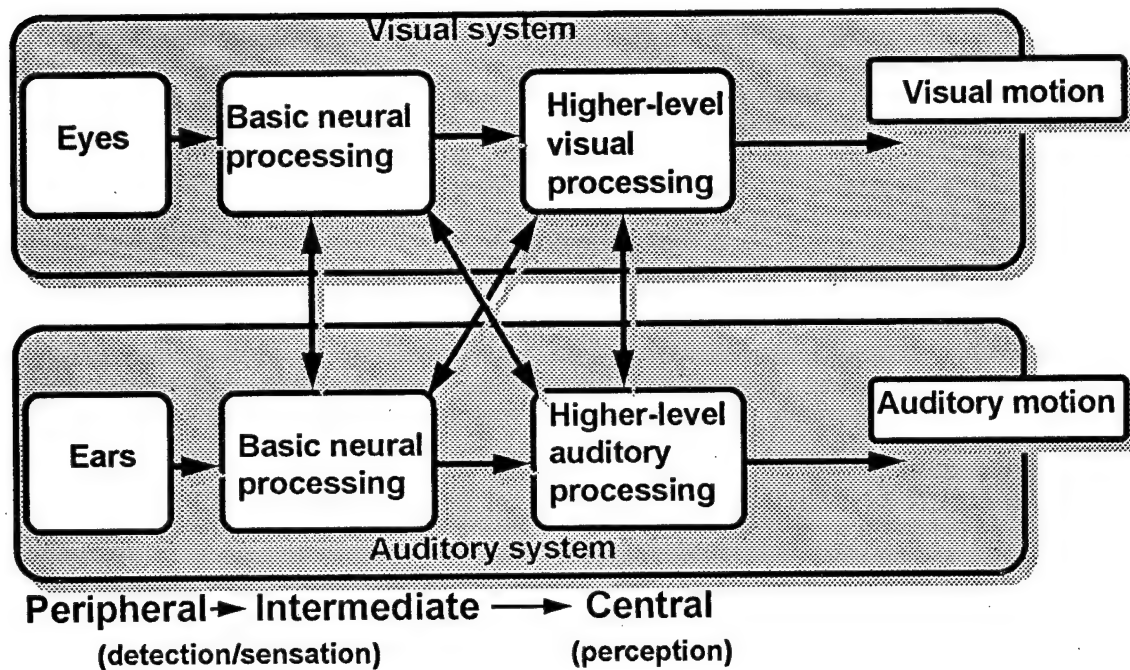


Figure 2.1: This simple overview model was composed of the end sense organs of vision and audition, early neural processing in both systems, late (or higher level) processing in both systems, and storage, or reporting, of resultant information.

## 2.2 Overview model of inter-sensory motion perception

A simple model of the auditory and visual sensory-perceptual system aids the formation of a framework supporting a literature review of inter-sensory motion perception as well as the organization of empirical results. The simple model can be thought of as an overview model in that it is non-quantitative but maintains a structure that supports the interactions of interest. The overview model is depicted in Figure 2.1.

This simple overview model is composed of the end sense organs of vision and audition, which are the eyes and ears, early neural processing in both systems, late (or higher level) processing in both systems, and storage or reporting of resultant information. The potential interconnections between these components are early processing to early processing, late processing to late processing, and early to late processing (or late to early processing). Linkages between the end organs themselves are not considered.

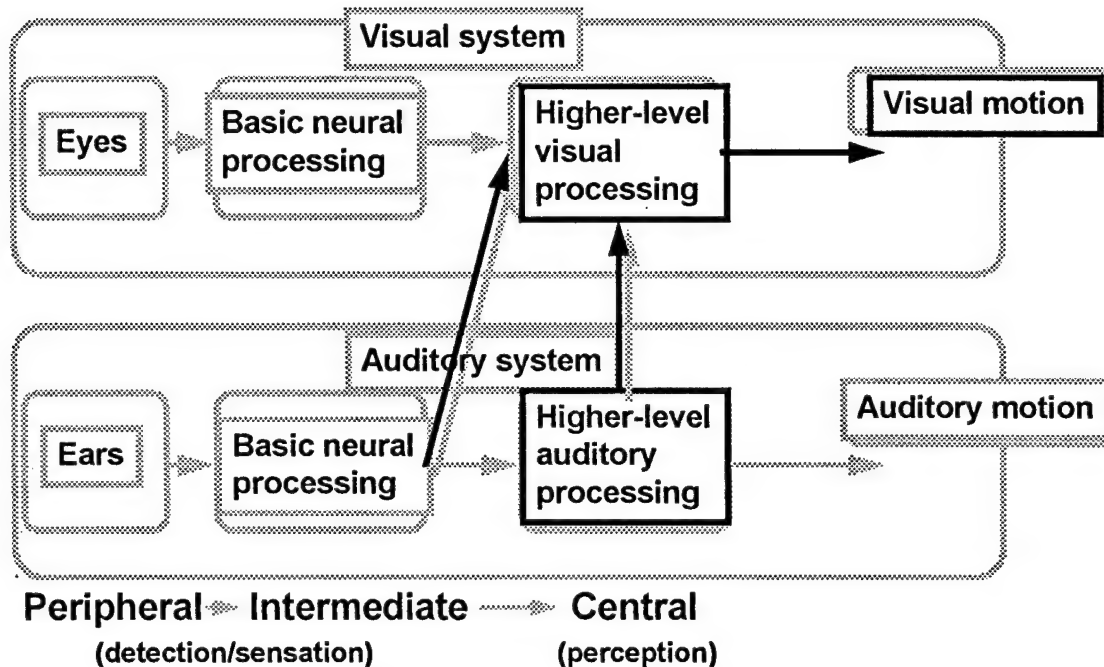


Figure 2.2: The focus of this research was the potential linkage between the early auditory processing and the late visual processing and the link between the late auditory processing and the late visual processing

Of the four potential links, the focus of this research is the potential linkage between the early auditory processing and the late visual processing and the potential linkage between the late auditory processing and the late visual processing. That focus is reflected in Figure 2.2 and is shown as darkened connecting lines.

## 2.3 Overview of the research

Eight experiments were performed investigating the spatial and temporal influence that moving auditory stimuli may exhibit over visual apparent motion perception.

Three experiments were performed investigating the influence of stroboscopic auditory information on the spatially-driven organization of a visual stroboscopic display.

Two experiments validated the use of a particular virtual auditory display device.

Number	Investigation area	Topic
1	Spatial Influence	Horizontal/vertical thresholds
2	Apparatus Validation	Velocity discrimination
3	Apparatus Validation	Minimum audible movement angle
4	Spatial Influence	Horizontal/vertical thresholds
5	Spatial Influence	Spatial Bias
6	Temporal Influence	Multi-stable temporal display
7	Temporal Influence	Temporal Bias
8	Manual control task	Intermittent tracking

Table 2.1: Temporal and spatial influence characteristics were investigated in eight experiments.

Once validated, the use of this device enabled better control over the movement of the stroboscopic auditory display reducing potential artifacts caused by spatially disparate aural and visual displays.

Two experiments were performed investigating influences of stroboscopic auditory information on the perceptual organization of temporally-driven visual stimuli.

A single experiment was performed evaluating human performance of an intermittent manual control tracking task using a visual-only target and a visual-plus-auditory target. The potential transition of auditory stimuli influences on the perception of simple visual displays into the more complex task of manual control was evaluated within this experiment.

The empirical topics and individual experiment titles are consolidated in Table 2.1 which provides an organizational structure describing the experiments detailed within this report.

## 2.4 Organization of the report

The report is organized into 10 chapters and appendixes. A brief description of each chapter's content is described in the following list.

Chapter 1: A general introduction to the research problem is described. Potential application areas are also discussed.

Chapter 2: The overall objectives of the research program are presented. A simple model of visual and auditory motion perception is proposed to highlight the focus of the research program. A brief description of the report organization is also included.

Chapter 3: A review of the literature that is pertinent to scientific investigation of the auditory influence on visual motion perception is presented. The topic areas include auditory and visual physiological structures involved in motion perception, existing mathematical models of visual and auditory motion perception, perceptual factors affecting visual and auditory motion perception. In addition, pertinent inter-modal literature regarding auditory and visual motion perception is reviewed in terms of physiological and perceptual factors. The complexity of the simple model proposed in Chapter 2 of this thesis is expanded into a mechanism model and an abstract model.

Chapter 4: A description of the apparatus utilized in the experiments described within this thesis is presented.

Chapter 5: A description and results of experiment one, an experiment investigating the characteristics of auditory influence on a spatially-driven visual percept of apparent motion, are presented.

Chapter 6: The methodology for an empirical validation of a virtual auditory stroboscopic display is developed and described. Results of the empirical validation, consisting of experiment two, an experiment investigating auditory velocity discrimination, and experiment three, an experiment investigating the minimum auditory movement angle of a virtual auditory stroboscopic display, are presented in this chapter.

Chapter 7: The description and results of experiment four, a second experiment investigating the characteristics of auditory influence on a spatially-driven visual percept of apparent motion using the virtual auditory display is presented. A comparison between results from the experiment described in Chapter 5 and this second evaluation is discussed. The description and results of experiment five, a third experiment evaluating characteristics of auditory influence on a spatially-driven visual percept of apparent motion, are also presented.

Chapter 8: The description and results of experiment six and seven, two experiments investigating the characteristics of auditory influence on a temporally-driven visual percept of apparent motion, are presented in this chapter.

Chapter 9: The transfer of inter-sensory perceptual influence into a complex task environment is described in this chapter through the results of experiment eight, an experiment involving manual pursuit tracking of an intermittent visual and auditory target.

Chapter 10: A summary of the experimental results and overall conclusions is presented. Modifications to the abstract model are presented which are based on conclusions from the experimental results. Recommendations for further empirical work investigating the auditory influence on visual motion perception are also described in this chapter.

Appendix A: A sample of the consent forms utilized in each of the experiments is contained in this appendix.

Appendix B: Instructions to the subjects for each of the visual/auditory experiments are contained in this appendix.

## Chapter 3

# Review of the literature

The literature review is structured around the composition and characteristics of a visual and auditory motion perception model founded upon information within the literature. Pertinent intra-sensory characteristics of motion perception are discussed in the review as are pertinent inter-sensory characteristics. Perceptual, physiological, and mathematical characteristics of motion perception are also discussed.

The research literature regarding motion perception is extensive. It is well beyond the scope of this review to describe all literature in the area of motion perception or even in the sub-areas of visual motion perception or auditory motion perception. Books continue to be written collating research findings in motion perception [5], [138]. As an alternative to critically reviewing all literature available regarding motion perception, only those topics that appear to have significant bearing on the influence of stroboscopic auditory stimuli on visual apparent motion perception are discussed.

### 3.1 Intra-modal literature

#### 3.1.1 Korte's law-based models of visual and auditory apparent motion perception

Research into visual motion perception has been performed for many years. The first modern study of motion was conducted by Sigmund Exner in 1875 [69]. Exner used the simplest stimulus for generating apparent motion, the flash of two light sources in succession, to first demonstrate apparent motion [36]. In this work, Exner demonstrated that humans could visually experience motion in an environment where they could not temporally or spatially resolve the sources of the motion [36].

Several characteristics of the visual stimulus could be manipulated by Exner within his experimental set-up. Spatial separation of the two sources, temporal duration of the sources, the duration between the cut-off of the first source and the turn-on of the second source, called the inter-stimulus interval (ISI), as well as, but



in a less controlled manner, the intensity of the visual source could be manipulated by Exner [36].

Exner concluded that three visual percepts could be elicited, those being simultaneous, sequential, and moving. Specifically, Exner found that at a temporal separation of 10ms the two sources appeared simultaneous, at longer temporal separations the two sources appeared as a single moving source, and at even longer separations the two sources appeared to be sequential [69]. These conclusions indicate that Exner found that spatial and temporal characteristics of visual stimuli had an impact on the visual perception of those stimuli. However, Kolers stated that Exner was less interested in stimulus characteristics required to elicit the perception of motion than in substantiating motion as a basic sensation, and because of this interest, did not pursue more extensive studies of the affect of stimulus characteristics on motion perception [69].

Max Wertheimer took Exner's ideas further by investigating in detail, how stimulus characteristics affected visual motion perception. Wertheimer reported a large number of visual apparent motion effects in a paper written in 1912 [69]. Wertheimer, like Exner, believed that motion was a basic sensation but Wertheimer found that Exner's three modes of visual perception did not have rigid boundaries in temporal separation [69]. Wertheimer also found that other visual percepts could be reliably elicited, such as broken motion perception [69].

A third scientist performing early research in apparent motion was Korte. In 1915, Korte published a set of *laws* which he developed from a systematic investigation of how spatial and temporal stimulus characteristics could be manipulated to maintain equivalent perceptions of visual apparent motion. *Korte's Laws* describe the relationships between four visual stimulus characteristics that must be maintained to elicit equivalent visual apparent motion perception. These *laws* provide some basic insight into the perception of apparent motion although they do not completely describe the perception of apparent motion. *Korte's Laws*, as taken from Boff [13], are shown below, where  $E$  is spatial distance, or extent, between stimulus presentations,  $I$  is the temporal interval between stimulus presentations,  $T$  is the duration of each stimulus presentation, and  $L$  is the luminance of the visual stimulus.

- 1  $E$  increases as  $L$  increases, with  $T$  and  $I$  held constant;
- 2  $E$  increases as  $I$  increases, with  $T$  and  $L$  held constant;
- 3  $L$  decreases as  $I$  increases, with  $T$  and  $E$  held constant;
- 4  $T$  decreases as  $I$  increases, with  $L$  and  $E$  held constant;

While *Korte's Laws* were derived from studies of visual motion perception, the earliest evidence of research concerning auditory apparent motion was Burt's article in 1917 demonstrating that apparent motion could be produced in the auditory modality [121]. In this article, Burt also demonstrated that the quality of the motion percept was a function of the ISI. In these studies the motion percept derived

from two sources of sound was reported as either simultaneous sounds, continuously moving, broken motion, or successive sounds, as the ISI was increased. Matheissen failed to elicit auditory motion reports in 1931 but Hisata replicated Burt's finding in 1934 [121].

It is clear from the work of these early researchers that humans experience the perception of motion both visually and aurally. It is also clear from Korte's very basic *laws*, specifically *law* number two, that the temporal and spatial characteristics of the visual stimulus affect the perception of apparent motion.

Further organization of the implications of stimulus spatial and temporal characteristics on apparent motion perception is achieved using a basic quantitative model derived from the overview model shown in Figure 2.2. Some aspects of this basic quantitative model can be built upon *Korte's Laws* augmented with several assumptions. The augmenting assumptions are:

- 1 The auditory and visual systems react similarly to temporal and spatial characteristics of apparent motion stimuli.
- 2 A single quantity can represent the robustness of the perceived motion within the human.
- 3 The two perceptual systems are not interconnected.

A graphic depiction of that model is shown as Figure 3.1. That model depicts characteristics of the visual and auditory stimulus, represented by  $I$  and  $E$  from *Korte's laws*, entering the sense receptors, either the eyes or the ears, on the left side of Figure 3.1, and perceptual characteristics reportable by the human observer, represented by  $C$ , exiting the model on the right side of Figure 3.1.  $I$ ,  $E$ , and  $C$  are measurable through the use of instrumentation or dialog with the human observer.

Within the model shown in Figure 3.1, two levels of processing are depicted, those levels being motion detection and motion perception. The variable depicted between these processing levels is the strength, or robustness, of the motion sensation. The strength of the motion sensation is represented by  $R$ .  $R$  can not be measured directly in the human but  $R$  is correlated with the characteristics of the stimulus and contributes to the perceptual and cognitive processes in the human necessary to respond to the stimulus. The subscripts  $a$  and  $v$  are subscripts on the variables  $C$  and  $R$  representing the auditory and visual modalities respectively.

The relationships depicted in Figure 3.1 are mathematically captured in the equations below, where  $R_v$  is the strength of the visual motion sensation and  $R_a$  is the strength of the auditory motion sensation.

$$R_v = f(E_v, I_v) \quad (3.1)$$

$$R_a = f(E_a, I_a) \quad (3.2)$$

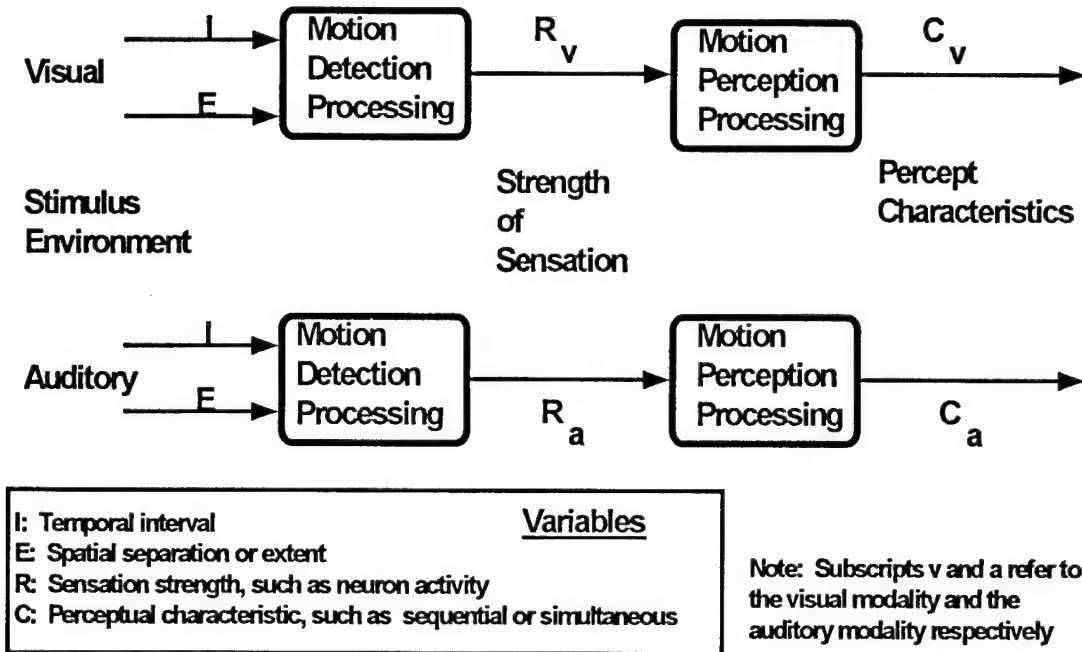


Figure 3.1: The temporal and spatial characteristics of the visual and auditory stimulus independently affect apparent motion perception.

and

$$C_v = f(R_v) \quad (3.3)$$

$$C_a = f(R_a) \quad (3.4)$$

Both  $R_v$  and  $R_a$  are functions of  $E$  and  $I$  in the visual and auditory environment.  $E$  is the spatial distance between stimulus presentations and  $I$  is the temporal duration between stimulus presentations.  $C_v$  represents the measurable response to visual stimuli and  $C_a$  represents the measurable response to auditory stimuli.

*Korte's Laws*, specifically law number two, expands the model quantitatively. Korte's second law presents a relationship between  $I$  and  $E$  to maintain a stable perception of motion when either variable is manipulated. However, Korte's laws do not specify what functional relationship exists between the strength of the sensation and  $E$  or  $I$ . Korte's laws also do not specify the relationship between the strength of sensation and percept characteristics. Korte simply states that to maintain a stable perception,  $E$  and  $I$  must be increased or decreased together.

This relationship can be captured in many forms mathematically. One such relationship is an additive relationship contained in the equations shown below. In these equations,  $\alpha$  and  $\beta$  are arbitrary constants. A graph of this relationship is shown as Figure 3.2.

$$R_v = \alpha_v E_v - \beta_v I_v + \text{constant} \quad (3.5)$$

$$R_a = \alpha_a E_a - \beta_a I_a + \text{constant} \quad (3.6)$$

Another potential mathematical representation is ratio-based. The ratio-based relationship, which also adheres to Korte's laws, is captured in the equations below. In these equations,  $\alpha$  and  $\beta$  are arbitrary constants but not necessarily equal to the constants in the previous equations. A graph of this relationship is shown as Figure 3.3.

$$R_v = \frac{\alpha_v E_v}{\beta_v I_v} \quad (3.7)$$

$$R_a = \frac{\alpha_a E_a}{\beta_a I_a} \quad (3.8)$$

*Korte's Laws* do not describe the boundaries of stimulus characteristics eliciting motion perception. Exner's original conclusions regarding the three perceptual modes, sequential, simultaneous, and moving, are not supported by these laws. As an example of Exner's conclusions within the context of this model, when presenting two spatially and temporally distinct sources, either visually or aurally, to a human,

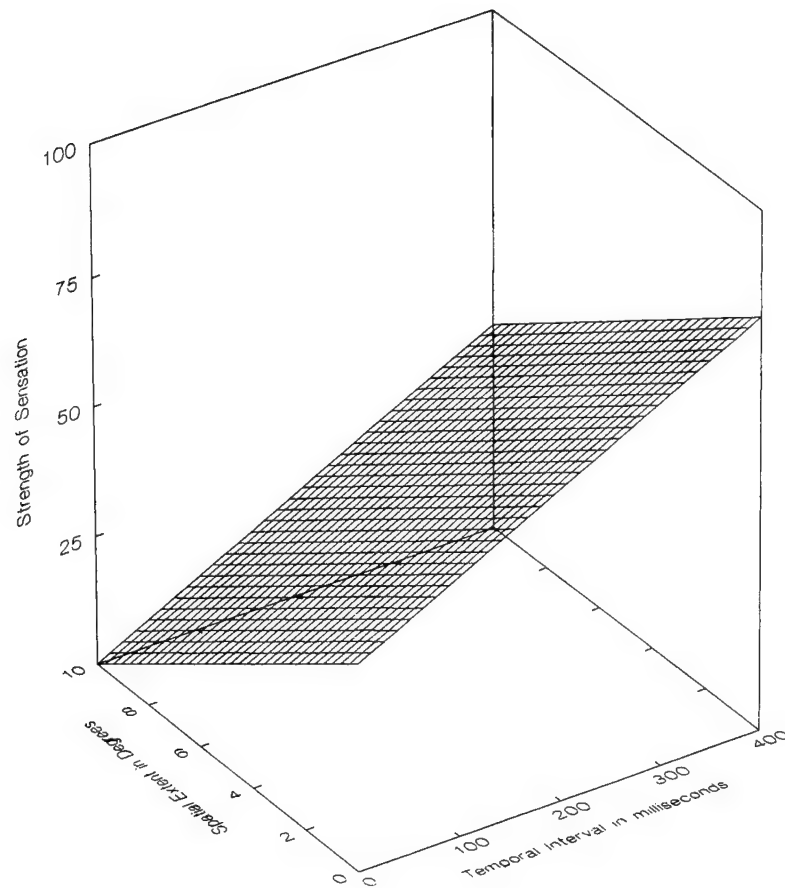


Figure 3.2: Korte's second law stated that to maintain a stable percept,  $E$  and  $I$  must be increased or decreased together. One such mathematical formulation is the linear combination of  $E$  and  $I$ .

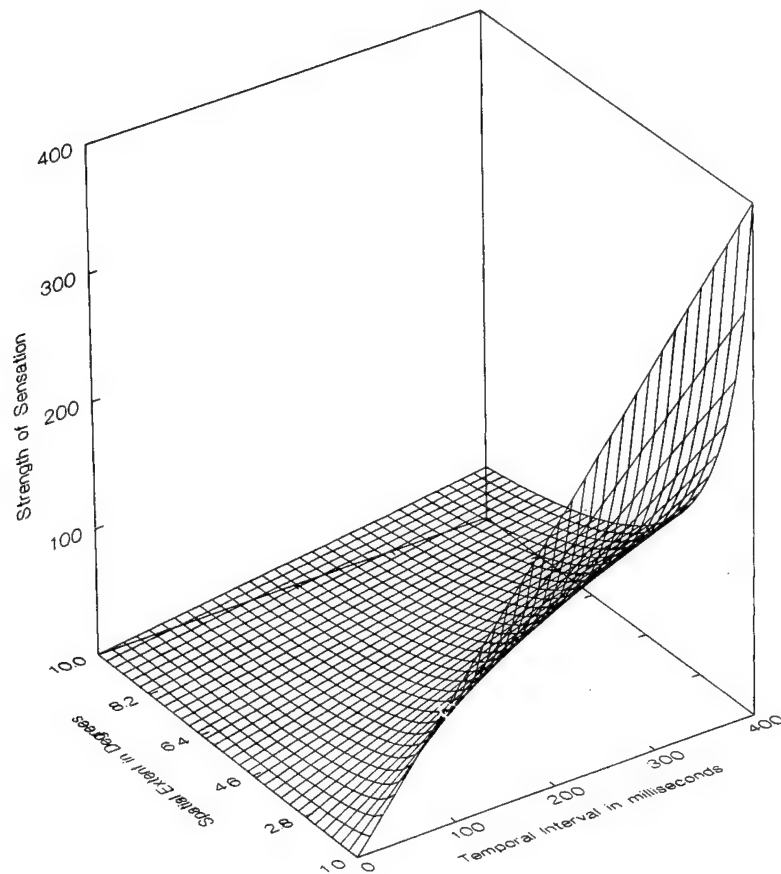


Figure 3.3: Korte's second law stated that to maintain a stable percept,  $E$  and  $I$  must be increased or decreased together. One such mathematical formulation is the ratio of  $E$  to  $I$ .

if the spatial separation of each stimulus presentation ( $E$ ) is too small, or the temporal interval ( $I$ ) is too small, the stimulus would be perceived and reported as two continuously presented dots of light or two continuously presented sound sources. If, on the other hand, the  $E$  is too large or the  $I$  is too large, the stimulus would be perceived and reported by the human as a series of successive illuminated dots or successively presented sound sources. This example begins to illuminate the complex relationship between strength of sensation and elicited perceptual characteristics, as well as required expansions of the quantitative model.

The boundary conditions of this example can be incorporated into a more complex model if an assumption is made regarding the strength of sensation. If the strength of sensation is a smooth, separable, function of both  $I$  and  $E$ , and the percept characteristics directly follow the strength of sensation, a graphic representation can be constructed. One such construction for  $I$  is shown as Figure 3.4 and a second construction for  $E$  is shown as Figure 3.5. These constructions depict a motion sensation strength that is a function of spatial and temporal factors.

The function relating the strength of motion sensation to both  $E$  and  $I$  is arbitrary in Figures 3.4 and 3.5 and is shown as the curve on the periphery of the shaded areas in these figures. In both motion sensation constructs, three categories of percept characteristics are associated with levels of  $E$  and  $I$ . The percept characteristics, simultaneous, sequential, and moving, are associated with arbitrarily levels of  $I$  and  $E$  on these graphs and represent the variable in  $C$  found in the model shown as Figure 3.1.

It is apparent from Figures 3.4 and 3.5 that  $C$  can not be derived only from the motion sensation strength. This is evident in that the simultaneous and sequential motion perception modes are associated with identical motion sensation strengths. Thus,  $R$  can only be thought of as a contributor to  $C$ .

The research following Korte's work has increased the knowledge of perceptual implications of stimulus characteristics and the basic quantitative model can be made more powerful by using this information. In general, many more factors have been implicated in motion perception than Korte described, and the overall understanding of the complexity of motion perception in humans has increased. Some of these relationships are discussed in the following sections.

### 3.1.2 Physiological structures

The physiology of the visual and the auditory system should be considered when investigating how stimuli affect perceptual characteristics and when constructing models which may functionally represent these systems. The physiology may provide insight into the potential complexity of the processes as well as highlight potential processing structures.

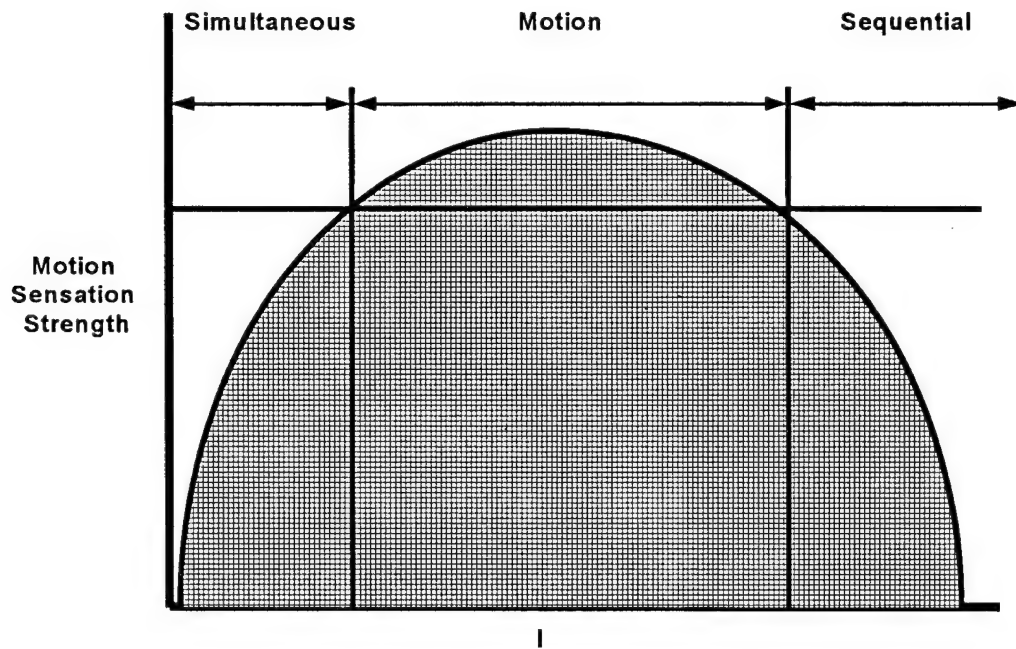


Figure 3.4: Motion sensation strength is a function of  $I$  and falls off at either large or small  $I$  values. The perception of motion may occur over a range of  $I$ .

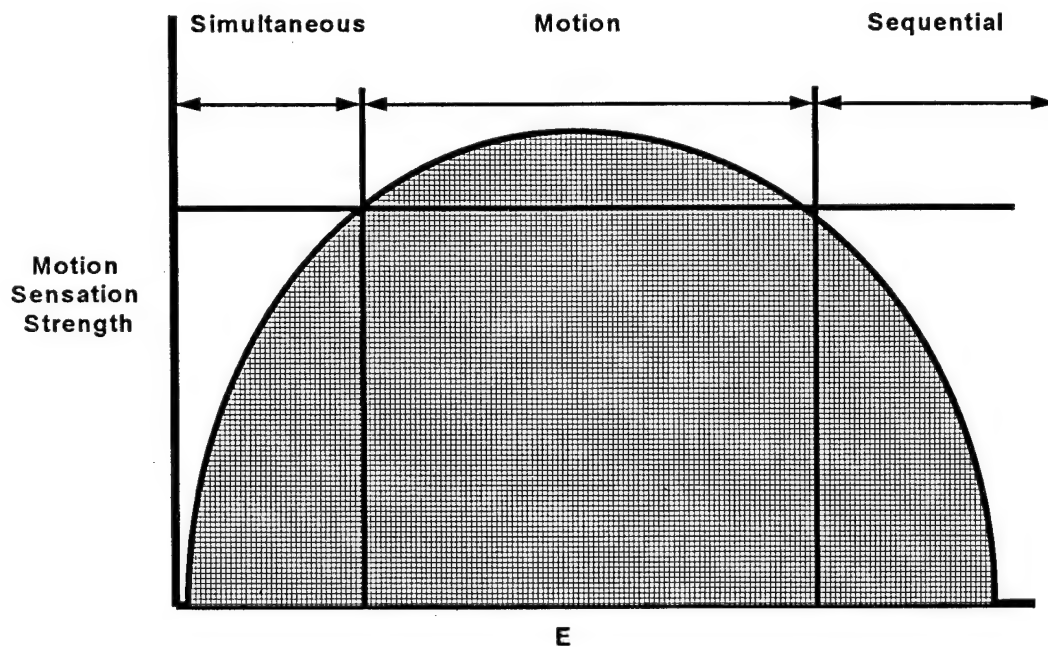


Figure 3.5: Motion sensation strength is a function of  $E$  and falls off at either large or small  $E$  values. The perception of motion may occur over a range of  $E$ .



### Visual physiology

Most of the literature regarding the physiological basis for visual motion perception that impacts this research attempts to describe physiologically-based mathematical or functional models of the mechanisms underlying motion perception. The visual physiology literature reviewed within this report is limited in scope and reflects the focus of this research, which is limited to intermediate-level and central-level processing. Most of this information comes from animal studies that may not perfectly correlate with the operation of the human visual system, but do propose possible architectures, functionality, and performance characteristics to be investigated in humans using other methods. As an example of this, Blasdel and Tootell traced the functional anatomy of the macaque monkey's sensitivity to moving visual stimuli and found that maps of orientation and sensitivity in the cortex differed between contoured and non-contoured moving stimuli [76].

The physiological pathway from the visual receptors to the brain in the human is complex. The receptor organs are the left and right eye, which incorporate an area of light sensitive cells in each eye called the retina. Each retina is connected via the optic nerve through a cross-over area, called the optic chiasm, to the lateral geniculate nucleus, or LGN. The LGN is connected to the visual area within the brain, called the visual cortex. The visual cortex was described by Spillman as being composed of several areas that were characterized by differing external connections, physiological properties, or topographic organization [111]. The visual system was also described by Spillman as being composed of many multiple paths of information flow, many of which were information specific [111].

A description of a possible motion-selective pathway within the visual system of the monkey was described by Lennie in a recently published book on visual perception [111]. The physiological elements involving visual system processing are shown in Figure 3.6, which were taken from a book by Graham [38], with the motion pathways highlighted [111].

Lennie postulated a motion-selective path that ran from the lateral geniculate nucleus of the thalamus (LGN) to the striate cortex (area V1), and on to the middle temporal area (MT), as well as an indirect path from the second visual area (V2) to MT and the third visual area (V3) to MT [111]. The strongest evidence linking this path to motion perception was that neurons in this path reacted distinctly to the motion of complex patterns [111]. This had been confirmed by Logothetis [74]. Graham implicated all of these areas as well as area MST in her description of motion-specific information flow paths [38]. Lennie stated that while the MT pathway may be tied to the analysis of motion, no knowledge existed concerning the contribution of the MT pathway toward motion perception. He also stated that the contributions made to MT, via V2 and V3, were not well understood [111]. One implication of this belief is that this pathway may incorporate some of the mechanisms of motion detection, which feed into the mechanisms supporting motion perception.

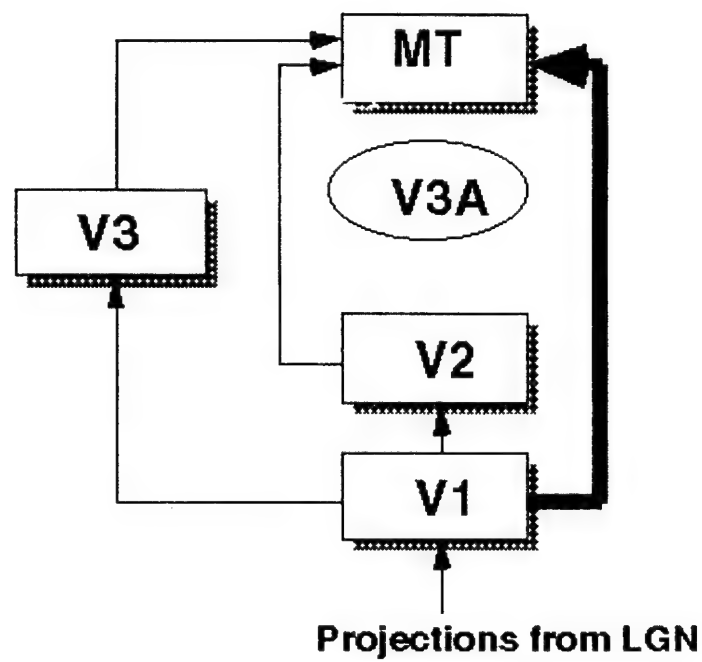


Figure 3.6: Many areas of the human physiology are involved in visual perception. Only a small portion of these areas is thought to be involved in visual motion perception. This construction is taken from Graham [38].

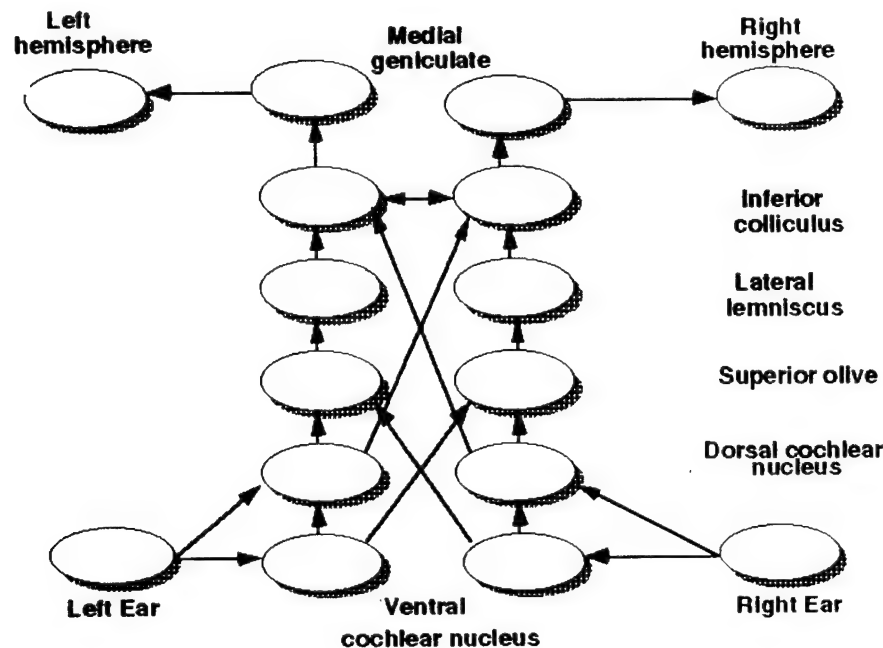


Figure 3.7: Many areas of the human physiology are involved in auditory perception. This construction is taken from Moore [83].

### Auditory physiology

The primary pathway of auditory information begins at the receptor organs, the left and right ear, and passes through the ventral cochlea nucleus, the dorsal cochlea nucleus, the superior olive, the lateral lemniscus, the inferior colliculus, the medial geniculate, and finally into the auditory cortex [83]. The elements within the auditory system are shown in Figure 3.7.

Much is known about auditory pathway physiology. However, little information was found in the literature concerning the physiological basis for auditory motion perception.

The most relevant information found in the literature regarding the physiology of auditory motion perception may be the physiological basis of auditory lateralization and localization. This information may have the same relevance to auditory motion perception that direction-tuned detectors in the visual system have to visual motion perception. Knudsen and Konishi found cells in the Barn Owl's mesencephalicus lateralis dorsalis, which according to Goldstein, is equivalent to the mammal inferior colliculus, that respond only to sounds originating within a small elliptical area in space [66] [65]. This small elliptical area could be considered the cells receptive field. Additionally, Knudsen and Konishi found that these cells were arranged to form a map of auditory space on the mesencephalicus lateralis dorsalis [66] [65]. Sovijarvi

and Hyvarinen found cells in the cortex of the cat which respond to sounds in specific areas in space as well as to sounds that move in specific directions [110].

Stern had utilized neurological data from the auditory nerve to develop a quantitative model of auditory lateralization [114]. Stern called this model the *position-variable* model. The model described subjective lateral position based on a 500Hz tone. While this was a detailed model that appeared to track closely with auditory nerve firing, it does not account for spectrally complex auditory stimuli nor does it predict perceptual characteristics.

Stern developed a second model, called the *weighted-image* model, which does not predict auditory nerve firing, but does begin to predict lateralization performance with spectrally complex stimuli [115]. This model was built around the concept of cross-correlation between the left and right auditory channels after band-pass filtering and rectification within each channel. The cross-correlation functions resulting from band-pass filtering each channel were compared for the consistent inter-aural delay across that particular band, as well as the magnitude of the delay across that band [115]. The predictive ability of the *weighted-image* model was further investigated by Trahiotis and Stern in 1989 where the model's performance was supported by two experiments [127].

Searle *et al* developed a statistical decision theory model of auditory localization based on non-linear regression of empirical results of 47 auditory localization investigations performed by 7 research teams [101]. Searle's model provides some degree of predictability regarding localization accuracy and some measure of the contributory weight of factors influencing auditory localization. Searle stated that the contributors for accurate auditory localization, ranked from largest to smallest, were inter-aural time delay, inter-aural head shadow, monaural head shadow, inter-aural pinna characteristics, monaural pinna characteristics, and shoulder bounce [101]. Searle did not, however, provide insight into the sensory, perceptual, or cognitive mechanisms involved in the localization of auditory stimuli.

### 3.1.3 Mathematical/computational models of visual motion detection and perception

The computational models for motion perception in the literature were developed as models of the human visual system. These computational models formulate visual motion perception as a temporal sequence of spatially distinct velocity vectors. The resultant sequence of two-dimensional vector arrays form the output of the computational models. The computational models predominantly model either the detection of motion, so called lower-level models, or concentrate primarily on higher-level aspects of motion perception, such as predicting human performance within a 3-dimensional structure from motion discrimination task.

A comprehensive survey and analysis of computational properties and networks concerning optical flow in biological and analog forms was presented by Poggio, Yang, and Torre in chapter 19 of Durbin's book on computing neurons [25]. One of the first

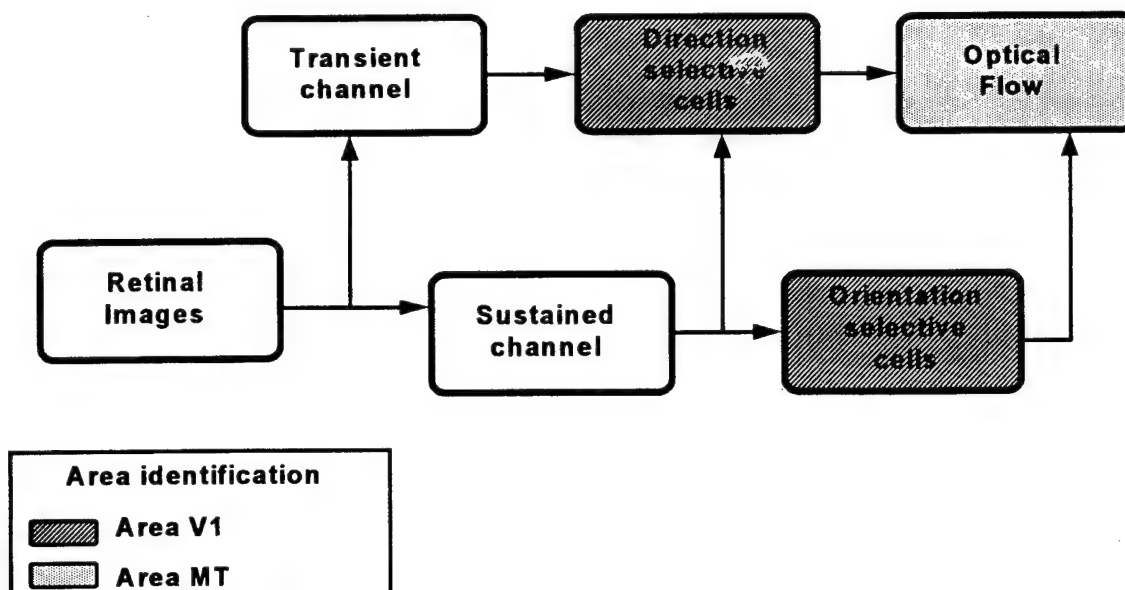


Figure 3.8: This computational model of neuronal activity involved in motion perception was taken from Wang [25].

points made in this work was that constraints imposed by the computational nature of motion perception and detection were not sufficient to exactly determine the implementation of the computational architecture that exists in humans [25]. Durbin also stated that the characteristics of the biological structures and characteristics in humans must be taken into account when determining which of the possible architectures may be employed to predict humans perceptual performance [25].

There were two steps necessary in the computation of optic flow according to Poggio *et al*, the first step was detection and the second was regularization [25]. Regularization was defined by Poggio as the use of a priori constraints to force the derived motion field vectors to be well behaved and to preserve discontinuities. The regularization step was required, according to Poggio, because the detection may be noisy, sparse, and/or non-unique [25].

A computational structure which correlates physiologic structures with algorithmic processes, taken from Durbin [25], is shown as Figure 3.8. The detection and regularization steps involving visual motion perception generally fit within the computational structure according to Wang [25].

### Motion detection

Several papers were published in the 1985 Journal of the Optical Society of America concerning computational models of visual motion detection [129][1][133]. This collection of papers had several themes that recurred through them. One of these

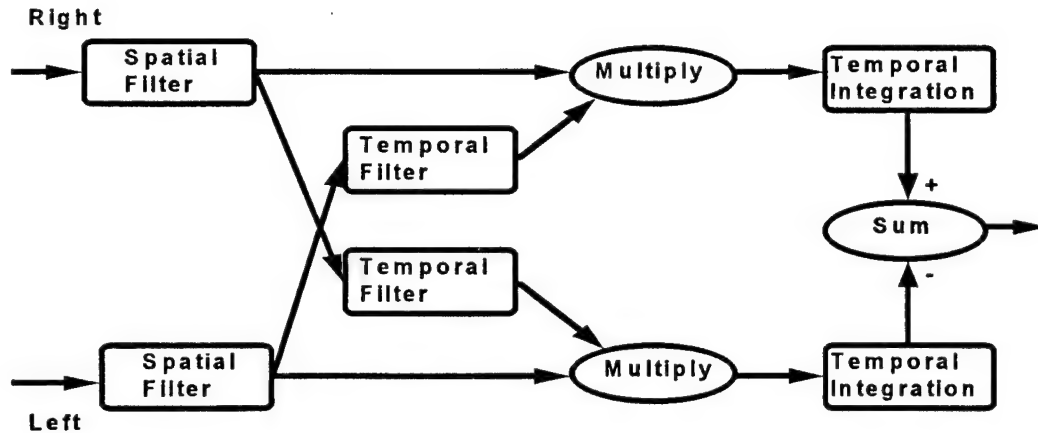


Figure 3.9: The Elaborated Reichardt detector model was a correlation model which incorporated multiplication.

themes was that the first computational model of human visual motion detection was developed by Reichardt and was first published in 1957 [129]. Spillmann agreed with this and stated that Werner Reichardt developed an elegant mathematical model of human motion perception in 1961 that provided the foundation for most existing mathematical models of human visual motion perception [111]. The Reichardt detector model utilizes a combination of linear time invariant temporal filters along with a multiplication stage and a subtraction stage. A Reichardt detector supplemented with a linear spatial filter was termed an Elaborated Reichardt Detector by Santen [129]. A diagram of the Elaborated Reichardt Detector model is shown in Figure 3.9 [129].

Several researchers have proposed computational models of the human motion detector that were functionally equivalent to the Elaborated Reichardt detector [129]. Adelson and Bergen developed a detection model that is shown in Figure 3.10 [1]. McKee also used this Adelson and Bergen model but modified it such that it would handle moving lines and points [78].

Watson and Ahumada developed a model that combined temporal and spatial filters into a main path and a quadrature path [133]. A block diagram of Watson's model is shown in Figure 3.11 [133]. The temporal filter in the main path consisted of the difference of two low pass filters. The spatial filter in the main path used a two-dimensional Gabor function, which consisted of an exponential multiplied by a cosine in the horizontal direction and an exponential in the vertical direction [133]. Equations representing the temporal and spatial filter models by Watson are shown below with  $t$  representing time,  $x$  and  $y$  representing the horizontal and vertical dimensions respectively.  $\lambda$  and  $u_s$  represent the *tuning* of the spatial filter such that  $\lambda = 3\frac{\sqrt{\ln 2}}{\pi} = 0.795/u_s$  and that the bandwidth is one octave [133]. The temporal filter is represented as  $T(t)$  and the spatial filter is represented as  $G(x, y)$  in the

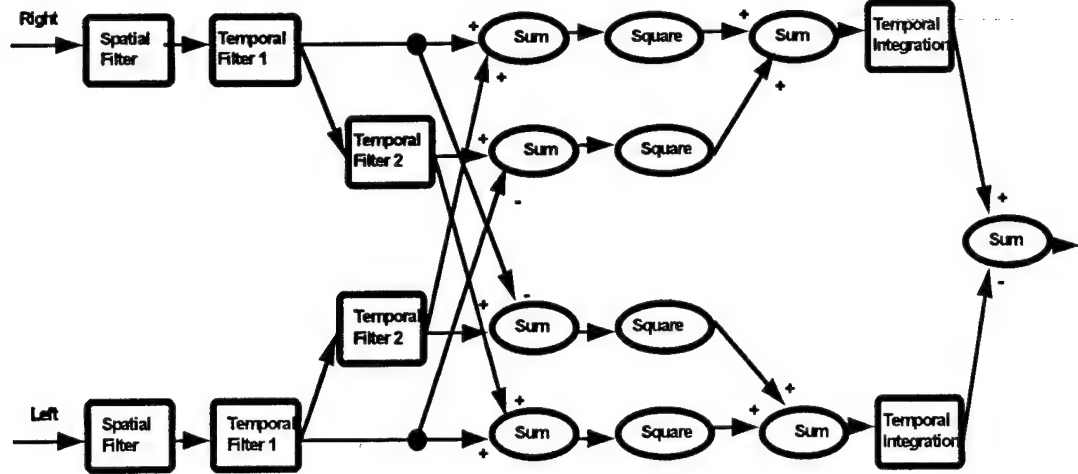


Figure 3.10: The Adelson and Bergen detector model was a correlation model which incorporated squaring, which may have been more plausible biologically than multiplication [25].

following equations.

$$T(t) = \xi \left[ \left[ \frac{u(t)}{(.004)8!} \left( \frac{t}{.004} \right)^8 e^{-t/.004} \right] - 0.9 \left[ \frac{u(t)}{(.0053)9!} \left( \frac{t}{.0053} \right)^9 e^{-t/.0053} \right] \right] \quad (3.9)$$

$$G(x, y) = e^{-(x/\lambda)^2} \cos(2\pi u_S x) e^{-(y/\lambda)^2} \quad (3.10)$$

Watson augmented the main path with a quadrature path incorporating spatial and temporal Hilbert transforms to generate directional characteristics [133].

Poggio *et al* described four motion detector types that directly resulted from the mathematical definition assumed for the optical flow [25]. There are vast differences in these detector models that appear to be directly associated with the underlying definition of optic flow for each detector. The detector models Poggio described are the Fennema-Thompson detector model, the Verri-Girosi-Torre detector model, the correlation-based detector models, and the shunting inhibition detector model. According to Poggio, correlation-based models may be a first order approximation to the biologically based architectures [25]. The shunting inhibition model was likely to be the scheme implemented by direction selective ganglion cells in the vertebrate retina [25]. It could be shown to be an approximation of the correlation model under certain conditions [25]. A block diagram of the shunting inhibition detector model, from [25], is shown in Figure 3.12.

Several neural network models had been constructed modeling the visual perception of motion. One model, constructed by Grossberg and Rudd [45], unified several other neural network models into a single model which predicted the perception of group and element motion of a 3-dot Ternus display described by Kolers [69]. The Grossberg model utilized a binocular version of the static boundary

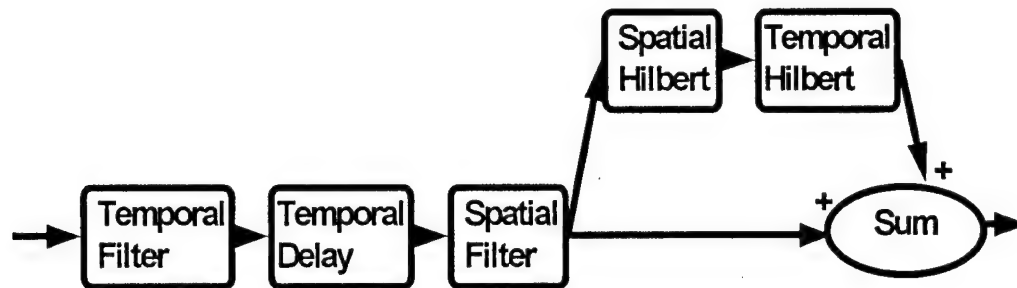


Figure 3.11: The Watson and Ahumada detector model incorporated a temporal filter, which consisted of the difference of two low pass filters, and a spatial filter, which consisted of a two-dimensional Gabor function.

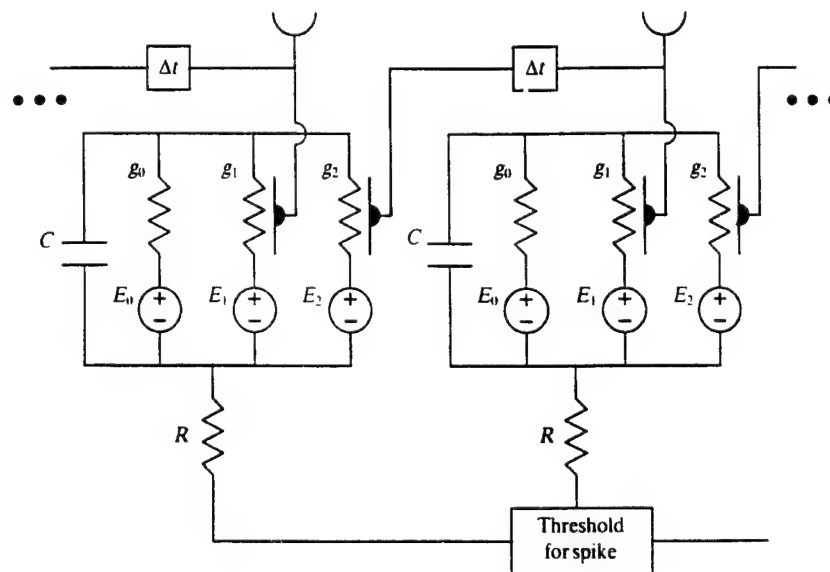


Figure 3.12: The shunting inhibition model was likely to be the scheme implemented by direction selective ganglion cells in the vertebrate retina [25].



contour system model [43] which contained an orientation-sensitive filter and a cooperative-competitive feedback loop. The orientation-sensitive filter modeled simple and complex cells of the V1 brain area with the cooperative-competitive feedback loop modeling higher area cells [45], [44]. According to Torre, the shunting inhibition network model may be considered an approximation to the correlation model under some conditions [126].

Hildreth organized detector models into only two categories, gradient-based and correlation-based [52]. Hildreth claimed that the correlation models appeared to be derived from early physiological studies of insects and resulted in the original Reichardt detector model [52]. Hildreth also claimed that the gradient models appeared to be developed from a mathematical perspective of image motion measurement [52]. Initial gradient models were proposed by Limb and Murphy with later models being developed by Marr and Hildreth [52]. Hildreth and Ullman proposed possible gradient-based algorithms for the measurement of visual motion orientation selective cells combined with mathematical implications for the biological computation of motion [53], [54]. Hildreth categorized motion algorithms that compare spatial and temporal derivatives of intensity as gradient schemes. She also described three specific computations for the measurement of visual motion involving combinations of convolutions with derivatives of Gaussians, temporal derivatives, and contour-specific computations [54].

According to Hildreth, psychophysical evidence could be obtained which supported the characteristics of either the gradient models or the correlation models [52]. For small contrast amplitudes within an image source, the gradient and correlation models were equivalent [52]. Hildreth stated that it may have been the case that the actual computations underlying motion perception in the human visual system had characteristics of both the gradient and correlation models at different biological levels within the central nervous system involved with motion perception [52].

### **Motion Regularization**

All of the previously described detector models required prior regularization of the input image data due to the explicit and/or implicit spatial derivatives that were part of the detector definitions [25]. In addition, the output of the detector models must be regularized, according to Poggio, because the detector model may produce noisy or locally sparse results [25]. Poggio stated that the regularization may take one of several forms, such as Gaussian blurring, minimization of a cost function, or Markov Random Fields [25].

As an example of the diversity of regularization solutions found in the literature, set of iterative equations computing two-dimensional optical flow fields was provided by Horn [56] which are in sharp contrast to the regularization algorithm developed by solving the detector equations using the Verri-Girosi-Torre optic flow definition [25].

Poggio claimed that most calculations within the detector and regularization models could be implemented in neuronal-based structures within the human visual

system [25]. Specifically, Poggio claimed that terms containing only partial derivatives in space could be implemented by groups of neurons which form elongated fields consisting of a central inhibitory area and two excitatory areas on opposing sides of the central area [25]. The direction of the elongation would then correspond to the particular spatial derivative being implemented.

Poggio also claimed that terms that had both spatial and temporal partial derivatives represented transient units and that both of these types of fields were present in the primary visual cortex of most mammals [25]. Poggio did concede, however, that some terms within these models, such as terms that required multiplication between transient fields and steady-state spatial fields, may not exist within mammalian visual systems [25].

Yuille and Grzywacz proposed a computational theory that linked a low-level motion detector in the human visual system with an algorithm that modeled characteristics of coherent visual motion at a higher processing level [143]. This algorithm utilized two stages, the measuring stage and the smoothing stage [143]. Yuille stated that any of the detector models may have been utilized with the computational model to describe the perception of coherent visual motion. Yuille proposed a cost function to be minimized as well as a solution to the minimization [143]. The solution is described by the equation below. In this equation, the spatial domain is represented by a two-dimensional set of points, represented by  $\vec{r}$ , the velocity at each point is represented by  $\vec{v}(\vec{r})$ , and  $i$  is an index over the entire field of  $\vec{r}$ .

$$\vec{v}(\vec{r}) = \sum_i \frac{\beta_i}{2\pi\sigma^2} e^{-\frac{|\vec{r}-\vec{r}_i|^2}{2\sigma^2}} \quad (3.11)$$

According to Yuille, the solution of velocity at any point in the spatial plane was a summation of velocities measured at surrounding locations that were weighted, as can be seen in the above equation in the form  $\beta_i$ , using a Gaussian curve based on the relative distance from the point being smoothed to the specific surrounding location, as can be seen in the term  $\vec{r} - \vec{r}_i$  [143].

### Summary of computational models of motion detection and perception

In summary, there are many computational structures and algorithms which can model some aspects of the human visual motion perception process documented in the literature. There is not a single, composite model which reconciles the different algorithms found within the literature. There are no computational models of auditory motion perception in the literature. In addition, no computational models were identified in the literature for visual and auditory motion system interactions.

### 3.1.4 Psychophysical factors affecting human motion perception

The research following Korte's work has produced several metrics of temporal separation which have been utilized throughout the literature. There appears to be two different measures of time which are typically used when describing the temporal separations of stroboscopic stimulus within the literature. The two measures are the inter-stimulus interval (ISI) and the inter-stimulus onset interval (ISOI). The ISI is measured between the end of one presentation and the start of the following presentation. The ISOI is measured between the start of one presentation to the start of the following presentation. ISOI is also called onset-to-onset (OTO) interval and stimulus onset asynchrony (SOA) in the literature. The ISOI must always be equal to or greater than zero. The ISI can be negative, zero, or positive.

The number and type of stimulus characteristics which are known to affect motion perception in humans has increased since Korte first published his perceptual laws. Some of these stimulus characteristics are described in the following sections.

#### The short-range process and the long-range process

There appears to be controversy in the literature concerning whether the processes governing the perception of apparent motion and real motion are identical. Anstis contrasted the relationship between real and apparent visual motion perception processes and concluded that there was evidence in the human visual system for two separate mechanisms for perceiving apparent motion, a short range process and a long range process [50]. Short range referred to small temporal or spatial shifts within visual stimuli, usually less than 10' to 15' of arc and less than 80ms to 100ms of inter-stimulus interval. Long range referred to larger temporal or spatial shifts [50] [54]. Braddick also described a two-component system which referred to short-range and a long-range components [14]. Anstis concluded that stroboscopic stimuli more complex than two flashing light sources were needed to begin to understand the relationship between apparent and real motion, and that complex displays also provided insight into the process sequencing of motion perception and form perception [50].

One of the most comprehensive articles found in the literature regarding the distinctions between the long-range process and the short-range process was written by Petersik [95]. In this article, Petersik argued that there were many visual percepts that could be attributed to different perceptual processes and the distinction between long-range and short-range apparent motion was useful [95].

Yuille modeled visual motion perception using a detection stage and a smoothing stage [143]. Yuille associated the short-range processing stage with motion detection, and associated the long-range processing stage with smoothing [143].

ISOI	Motion Perception
< 30ms	Simultaneous
30ms to 60ms	Continuous Movement
$\approx$ 60ms	Optimal Movement
60ms to 200ms	Broken Movement
> 200ms to 300ms	Successive

Table 3.1: Visual apparent motion could be perceived over a range of temporal intervals but the characteristics of the perceived motion differed

### Spatial and temporal effects within visual stimuli

In the early 1930s, Neuhaus found that visual motion perception could be elicited using dots separated by  $0.5^\circ$  and ISIs ranging from 50ms to 250ms and at  $4^\circ$  separations the ISI could lie between 100ms and 160ms [12]. In the early 1950s, Zeeman was able to obtain reported visual motion perception across  $2^\circ$  to  $18^\circ$  separations [144].

As a further example of the range of motion perception quality that can exist, Table 3.1, which was taken from Goldstein, depicts characteristic ISOIs of a two-light-source stimulus and places the visual apparent motion perceptions into 5 categories when all other factors were held constant. These perceptual categories are utilized throughout the visual apparent motion literature.

Sekuler found that training could affect visual motion discrimination and developed a model of motion perception that utilized broadly tuned direction-sensitive detectors [103], [102]. Ball determined in a study in 1985 that the directionally-tuned motion detectors were separated by at least  $120^\circ$  to  $150^\circ$ . This study addressed direction assessment and not simple motion detection [6].

McKee found that the stimulus duration time needed to discriminate motion direction for a single visual target was approximately 100ms. McKee also claimed that for stroboscopic stimuli, the strobe rate should be above 10Hz. Additionally, McKee found that a 3 cycle per degree grating produced motion discrimination if the grating velocity was kept below  $1^\circ \text{sec}^{-1}$  [79].

Kolers provided a comprehensive account of the contributions of several early researchers in the area of visual apparent motion [69]. He compared some of the results of these researchers in terms of the characteristics required to elicit the perception of apparent motion as functions of spatial separation and ISOI [69].

Petersik and Rosner investigated the effect of position cues on the perception of a bi-stable visual display [97]. The bi-stable display used was a modification of a Ternus display with bars connecting two of the three dots in each frame. The connecting bars would either remain fixed with the display or move as the group of dots moved. The temporal frequency at which this type of display alternated caused

the viewer to perceive either *group motion* at large ISIs or *element motion* at smaller ISIs [96]. Petersik found that the connecting bars would affect the visual perception such that when the bars moved with the dots from frame to frame, reports of *group motion* would increase and when the bars remained fixed from frame to frame, reports of *element motion* would increase [97]. This result suggested that additional intra-modal cues, that were linked to the Ternus display, aided the formation of the perception by providing information substantiating either *group motion* or *element motion* perception of the Ternus display [97].

Breitmeyer and Ritter investigated the perception of a Ternus display at a between-element spatial extent of  $1.2^\circ$ , frame durations of 50ms and 200ms, and ISIs ranging between 10ms and 100ms [16]. Breitmeyer found that the percentage of group motion reports increased as the frame duration increased and as ISI increased [16]. Breitmeyer also discussed the effect of visual persistence on the transition from group to element motion of Ternus displays when the ISI was manipulated as an independent variable. Breitmeyer stated that using a modified version of the method of constant stimuli would reduce the potential for selective adaptation reported by Petersik and Pantle [96] when using a method of ascending and descending limits [16].

In the literature, Strybel documented a comparison of apparent motion perception reports from stroboscopic auditory and visual stimuli [120] [123]. Strybel experimentally manipulated stimuli horizontal extent and the inter-stimulus interval to obtain reports of the quality of the apparent motion perception. Other researchers such as Burt and Sperling, have also captured quantitative relationships between the strength of the apparent motion perception and the spatial and temporal stimulus characteristics [17]. The internal variables of the model in Figure 3.1,  $R$  and  $C$ , were derived from the work of Strybel and others, and are described in the following sections.

Mathematically, the strength of the motion sensation, represented by the variable  $R$ , is a function,  $f$ , of both space and time. Space is measured as the spatial extent of the stimulus and represented as the variable  $E$ . Time is measured as the temporal duration of the ISOI and represented by the variable  $I$ .

$$R = f(E, I) \quad (3.12)$$

The percept characteristics are, in turn, a function of the strength of the sensation as shown in the equation below.

$$C = f(R) \quad (3.13)$$

When the function relating the sensation strength and the temporal and spatial variables is specified, quantitative prediction of human perceptual characteristics is possible. One such function relating spatial and temporal qualities to sensation strength was constructed by Burt and Sperling [17] using spatial extents associated

with the short-range process between 3' and 29' of arc and eliciting perceptual strength reports from subjects. Burt and Sperling formulated an exponential-based model of the form shown below [17]. In these equations,  $I$  is the ISOI of the display and  $E$  is the element separation distance.

$$C_v = F_t(I)F_d(E) \quad (3.14)$$

where

$$F_t(I) = I^{\alpha\beta} e^{-\beta I} \quad (3.15)$$

and

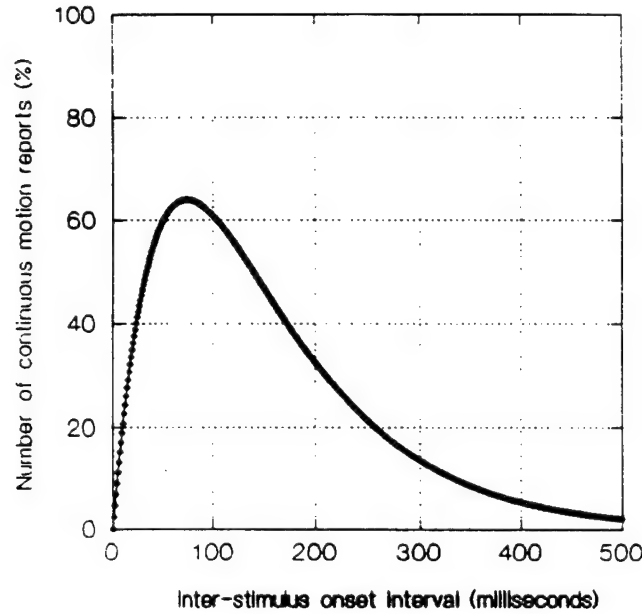
$$F_d(E) = (e^{-\gamma/E})/E \quad (3.16)$$

Yuille modeled the effect that spatial separation had on the strength of visual motion sensation in a different way than did Burt and Sperling. Yuille modeled this effect using a two-dimensional array of Gaussians centered about each motion detector [143]. The Watson model utilized two-dimensional arrays of spatial and temporal linear filters in combination with Hilbert transforms enabling directionality of each motion detector. Watson used as a temporal function the low-pass filter originally derived by Fourtes and Hodgkin and a spatial filter described by Sakitt and Barlow [133].

The data from Strybel, obtained using a constant duration of 50ms [120] [123], were used in this report to form a set of exponential functions which could be associated with the long-range process. Taking a parsimonious approach and assuming the functions of space and time were separable, a simple quantitative representation was constructed using exponentials.

The onset of the exponential models the rapid onset of motion perception from short spatial separations and short ISOIs. The back edge of the exponential models the ability to maintain a perception, with slight dissipation of strength, over greater separations and ISOIs. The visual components of the percept characteristic variable are represented by  $C_v$ . The subscript of the percept characteristic variable indicates  $C$  for the continuous motion percept and  $B$  for the broken motion percept. In this way,  $C_{vC}$  is the percentage of continuous motion reports in the visual modality,  $C_{vB}$  is the percentage of broken motion reports in the visual modality.  $E$  is in degrees measured horizontally and  $I$  is the ISOI in milliseconds in these equations with  $v$  representing the visual modality.

$$C_{vC}(E_v, I_v) = \cos(0.012E_v)(-300.0e^{-0.018I_v} + 300.0e^{-0.01I_v}) \quad (3.17)$$



**Figure 3.13:** The percentage of continuous visual motion reports can be represented as a function of the inter-stimulus onset interval if the duration is held constant. The equations representing this data were developed within this report based on data taken from Strybel [120] [123] at a duration of 50ms and an extent of  $10^\circ$ .

$$C_{vB}(E_v, I_v) = \cos(0.012E_v) \left( \frac{I_v}{2.17} \right) e^{-\left( \frac{I_v}{300.6} \right)^2} \quad (3.18)$$

Graphs, based on these equations, of percept characteristic strength at a single spatial extent are shown in Figures 3.13 and 3.14. Three dimensional plots of perceptual strength as functions of both spatial extent and ISOI are shown in Figures 3.15 and 3.16.

### Spatial and temporal auditory stimulus effects

The definition of *auditory* apparent motion perception is roughly equivalent to the definition of *visual* apparent motion perception. However, several subtle implications arise from the differences between the sense organs themselves as well as neuro-physiological structures and organization differences between the visual and auditory modalities.

The minimum audible movement angle (MAMA) is a measure of dynamic resolution of the auditory system. The MAMA is the dynamic equivalent of the minimum audible angle (MAA), which is the smallest perceivable change in auditory source position. The MAMA is defined as the spatial amplitude required to

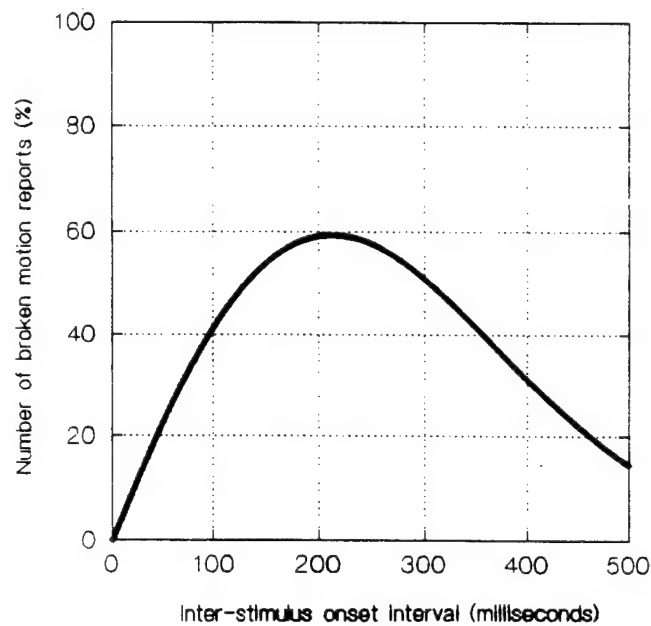


Figure 3.14: The percentage of visual broken motion reports can be represented as a function of the inter-stimulus onset interval if the duration and extent are held constant. The equations representing this data were developed within this report based on data taken from Strybel [120] [123] at a duration of 50ms and an extent of  $10^\circ$ .



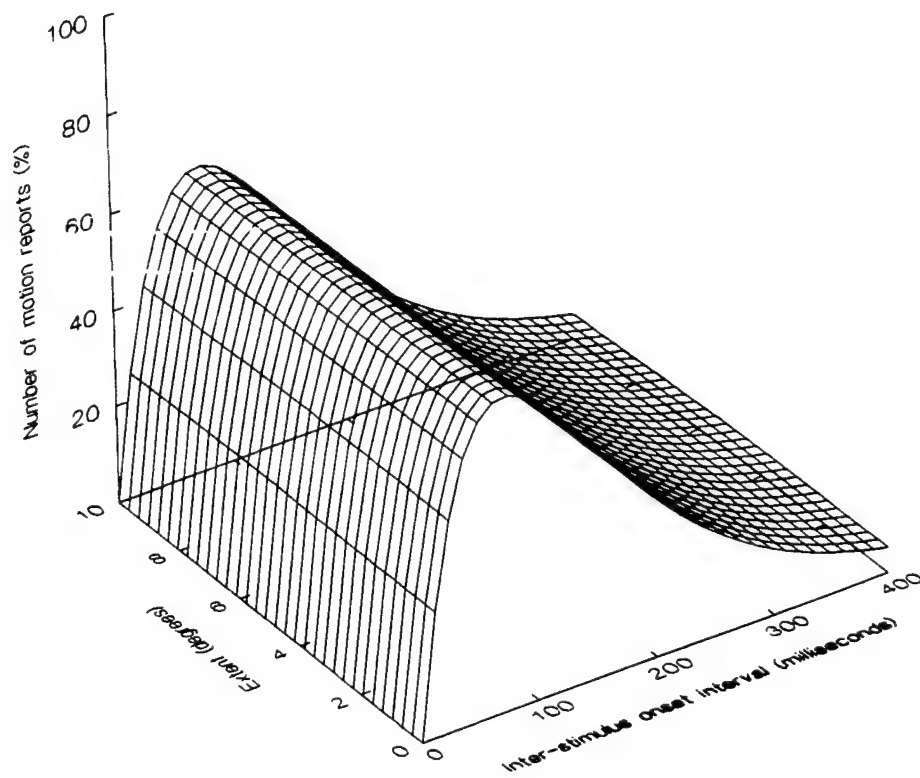


Figure 3.15: The representation in Figure 3.13 was expanded to include extent as a variable using data from Strybel [120] [123] at a constant duration of 50ms.

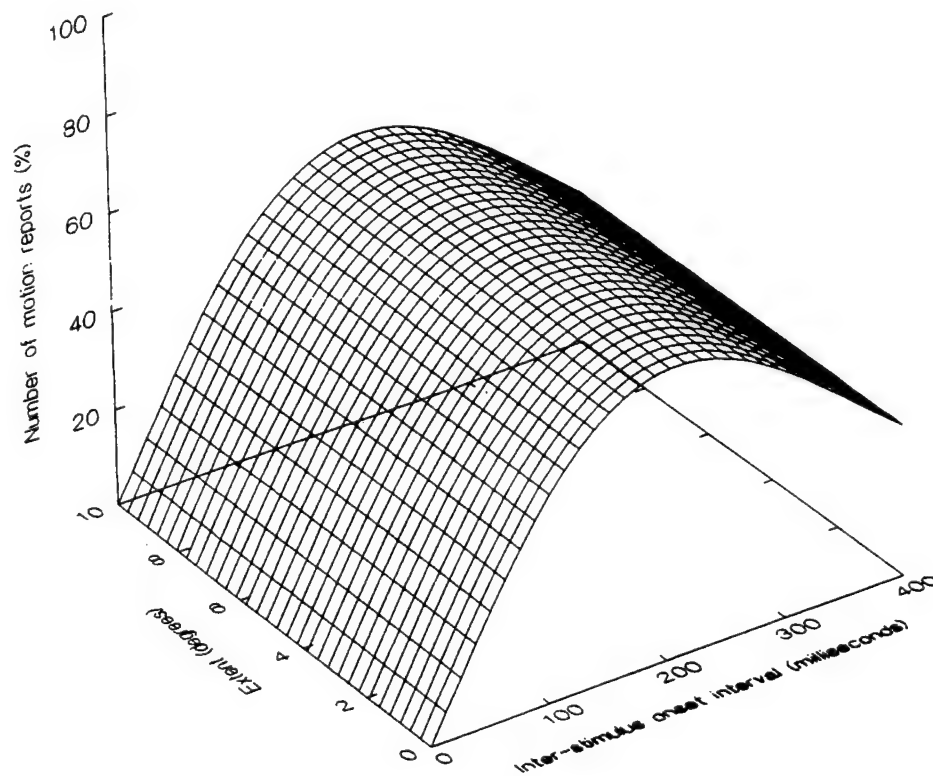


Figure 3.16: The representation in Figure 3.14 was expanded to include extent as a variable using data from Strybel [120] [123] at a constant duration of 50ms.

distinguish moving from stationary auditory sources. Perrott found the MAMA to be a linear function of auditory source speed when evaluated using a 500Hz tone between  $90^\circ \text{sec}^{-1}$  and  $360^\circ \text{sec}^{-1}$  [93]. He found that the MAMA was approximately  $8.3^\circ$  at  $90^\circ$  per second and approximately  $12.9^\circ$  at  $180^\circ \text{sec}^{-1}$  [93]. Perrott argued that the linear relationship could not be totally an artifact of duration changes resulting from higher velocities of the sound source [93]. Perrott also stated that presentation timing of at least 300 milliseconds must be used to ensure velocity perception [34].

Strybel, Manligas, and Perrott determined that the MAMA was also a function of sound source location relative to the listener's head using a broad-band audio signal (500Hz-8000Hz) moving at  $20^\circ \text{sec}^{-1}$  [122]. These researchers found that the mean minimum MAMA, approximately  $1.1^\circ$ , was obtained in front of the subject and increased to approximately  $3.4^\circ$  at  $80^\circ$  to the right and left horizontally [122]. When the elevation of the auditory source was manipulated, the mean minimum MAMA, approximately  $1.1^\circ$ , that was obtained in front of the subject, increased to approximately  $2.5^\circ$  at  $85^\circ$  in elevation [122]. Analyses of variance indicated that the mean MAMA did not change as a function of location except at the  $80^\circ$  horizontal locations and the  $85^\circ$  vertical location [122].

Harris and Sergeant investigated the contribution of frequency content on the dynamic spatial resolution, as measured by the MAMA, of auditory perception in 1971 [49]. They found that the frequency content of the auditory source did affect the MAMA but employed only three subjects [49]. Specifically, Harris and Sergeant found that the MAMA ranged from  $1.2^\circ$  to  $1.6^\circ$  at  $2.8^\circ/\text{sec}$  using an 800Hz tone, a doubling of the MAMA at 1600Hz and, a MAMA of  $0.6^\circ$  to  $0.9^\circ$  at  $2.8^\circ/\text{sec}$  using an 3200Hz tone [49]. However, not all three of the subjects reflected this pattern [49].

Harris reported a significantly different result in 1972. Harris found that the MAMA ranged from  $2.5^\circ$  to  $4.4^\circ$  at  $2.8^\circ/\text{sec}$  and  $2.5^\circ/\text{sec}$  using an 800Hz tone, and no difference using 800Hz or 1600Hz tones [48]. In this study, Harris again found that not all subjects reflected the same patterns for MAMA shifts due to frequency content of the auditory source [48]. Harris found no systematic effect of auditory source frequency on MAMA [48].

Perrott and Tucker investigated the effect of both auditory source frequency content and velocity on the MAMA [94]. They found for auditory sources below 1000Hz, the MAMA was smaller than for auditory sources above 1000Hz [94]. They also found that the MAMA increased as the velocity of the auditory source was increased and that the shape of the MAMA curves at each velocity evaluated, between  $8^\circ/\text{sec}$  and  $128^\circ/\text{sec}$ , corresponded favorably with the shape of static auditory spatial resolution curves plotted as a function of auditory source frequency [94].

Perrott stated that the effect of frequency content on MAMA, operating in two bands (above and below 1000Hz), substantiated the belief that the same process that mediated auditory spatial acuity with static sources also mediated auditory spatial acuity with dynamic sources [94]. This process, as described by Moore and many other researchers, consisted of one mechanism( such as inter-aural phase differences)

ISOI	Motion Perception
< 20ms	Simultaneous
20ms - 110ms	Continuous Movement
110ms	Optimal Movement
110ms - 200ms	Broken Movement
> 200ms	Successive

Table 3.2: Auditory apparent motion could be perceived over a range of temporal intervals but the characteristics of the perceived motion differed

mediating localization of lower frequency tones, a separate mechanism (such as inter-aural intensity differences) mediating localization at higher frequency tones, and for middle-frequency tones, neither mechanism operating as effectively leading to minimum spatial acuity for the middle-frequency ranges [83].

Lakatos investigated the critical SOA at which successive tones presented at different locations could not be differentiated. He investigated spatial extents of auditory stimuli between 20° and 110° horizontally and between 25° and 67° vertically. Lakatos found that the critical SOA did increase as the spatial extent increased, ranging from approximately 100ms at 20° to 200ms at 110° [73]. He found that vertical separation, over the areas evaluated, did not affect the critical SOA [73].

Perrott discussed the relationship between the perception of auditory apparent motion and the duration of pulsed stimuli for pulses ranging from 10ms to 300ms [34]. On the basis of Perrott's findings, a table could be constructed which described how auditory motion perception, as classified in 5 categories, was affected by inter-stimulus onset intervals (ISOI) when all other factors were held constant. This table is shown as Table 3.2.

Strybel reviewed the literature regarding the perception of real, simulated, and apparent auditory motion. Strybel stated that the most obvious result from his review was the need for more information regarding auditory motion perception [119]. Strybel stated that frequency affected the MAMA of real sound sources such that the MAMA for signals below 1000Hz was smaller than for frequencies above 1000Hz and that the largest MAMAs were found between 1200Hz and 2000Hz [119]. The sensitivity of the auditory system to stimulus displacements was poor compared to the visual system but the ability to judge stimulus velocity was comparable in both modalities [119]. Strybel also stated that auditory apparent motion was a robust phenomenon and that monaural information could be used alone to generate auditory apparent motion without the ability to perceive direction [119].

Briggs and Perrott in 1972 showed that as the duration of an auditory stimulus increased, the ISI that produced optimal movement reports decreased [121]. Perrott demonstrated that neither rise-time or correlation within auditory stimuli affected the ISI that produced apparent motion [121].

In the literature, Strybel documented a comparison of apparent motion perception reports from stroboscopic auditory and visual stimuli [120] [123]. Strybel experimentally manipulated the horizontal extent and the inter-stimulus interval of stimuli to obtain reports of the quality of the apparent motion perception.

The data from Strybel [120], [123] were used to form a set of exponential functions which could be associated with the long-range process for auditory apparent motion perception. The Strybel data were obtained using a constant duration of 50ms. Taking a parsimonious approach, assuming the functions of space and time were separable and using 50ms duration data, a simple quantitative representation was constructed using exponentials.

The onset of the exponential models the rapid onset of motion perception from short spatial separations and short ISOIs. The back edge of the exponential models the ability to maintain a perception, with slight dissipation of strength, over greater separations and ISOIs. The auditory components of the percept characteristic variable are represented by  $C_a$ . The subscript of the percept characteristic variable indicates  $C$  for the continuous motion percept and  $B$  for the broken motion percept. In this way,  $C_{aC}$  is the percentage of continuous motion reports in the auditory modality, and  $C_{aB}$  is the percentage of broken motion reports in the auditory modality.  $E$  is in degrees measured horizontally and  $I$  is the ISOI in milliseconds in these equations with  $a$  indicating the auditory modality.

$$C_{aC}(E_a, I_a) = \cos(0.012E_a)95.0e^{-\left(\frac{I_a-40.0}{25.0}\right)^2} \quad (3.19)$$

$$C_{aB}(E_a, I_a) = \cos(0.012E_a)(-190.0e^{-0.018I_a} + 190.0e^{-0.005I_a}) \quad (3.20)$$

Graphs, based on these equations, of percept characteristic strength at a single spatial extent are shown in Figures 3.17 and 3.18. Three dimensional plots of perceptual strength as functions of both spatial extent and ISOI are shown in Figures 3.19 and 3.20.

One characteristic of these equations clearly seen by these plots, and the similar equations and plots derived from the Strybel data in the visual domain, is that the ISOI eliciting the optimal perceptual strength does not change with spatial extent over the range of ISOI and extent utilized. Another characteristic of these equations is that the ISOI range eliciting any specific strength decreases as spatial extent increases. This characteristic was also depicted in empirical data from Kolers [69] and Burt and Sperling [17]. These characteristics indicate that the ISOI and the extent can be modeled as separable functions.

Another interesting attribute of these equations is the resemblance of the graph relating continuous motion reports to the low-pass filter gain utilized by Watson [133]. If the ISOI is doubled and inverted, it can be viewed as a very crude approximation to frequency, measured in cycles  $\text{sec}^{-1}$ . As can be seen in Figure 3.13, optimal reporting percentages occurred at an ISOI of approximately 70ms which

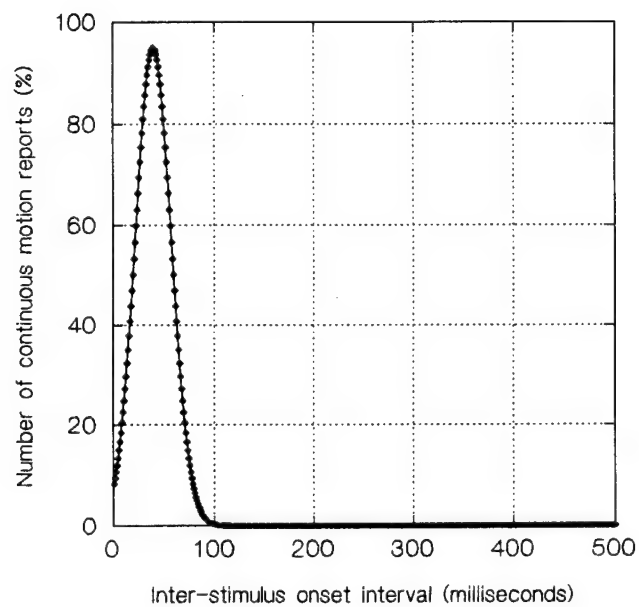


Figure 3.17: The percentage of continuous motion reports can be represented as a function of the auditory inter-stimulus onset interval if the duration and extent are held constant. The equations representing this data were developed within this report based on data taken from Strybel [120] [123] at a duration of 50ms and an extent of  $10^\circ$ .

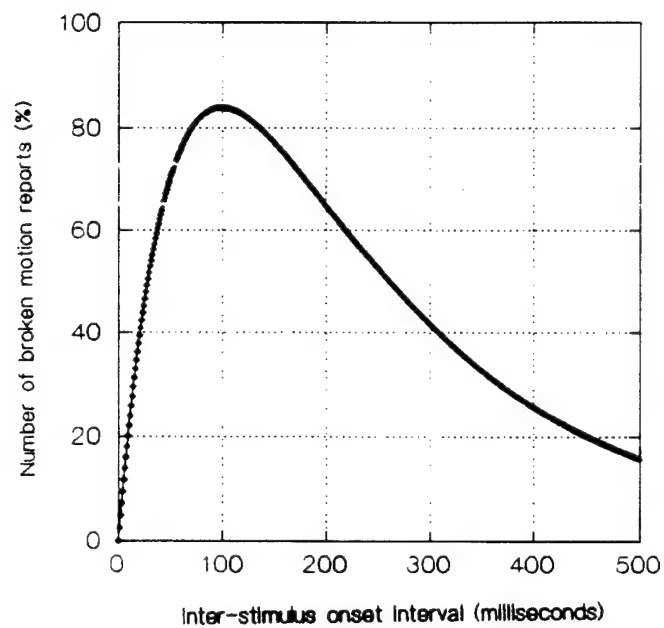


Figure 3.18: The percentage of broken motion reports can be represented as a function of the auditory inter-stimulus onset interval if the duration and extent are held constant. The equations representing this data were developed within this report based on data taken from Strybel [120] [123] at a duration of 50ms and an extent of  $10^\circ$ .

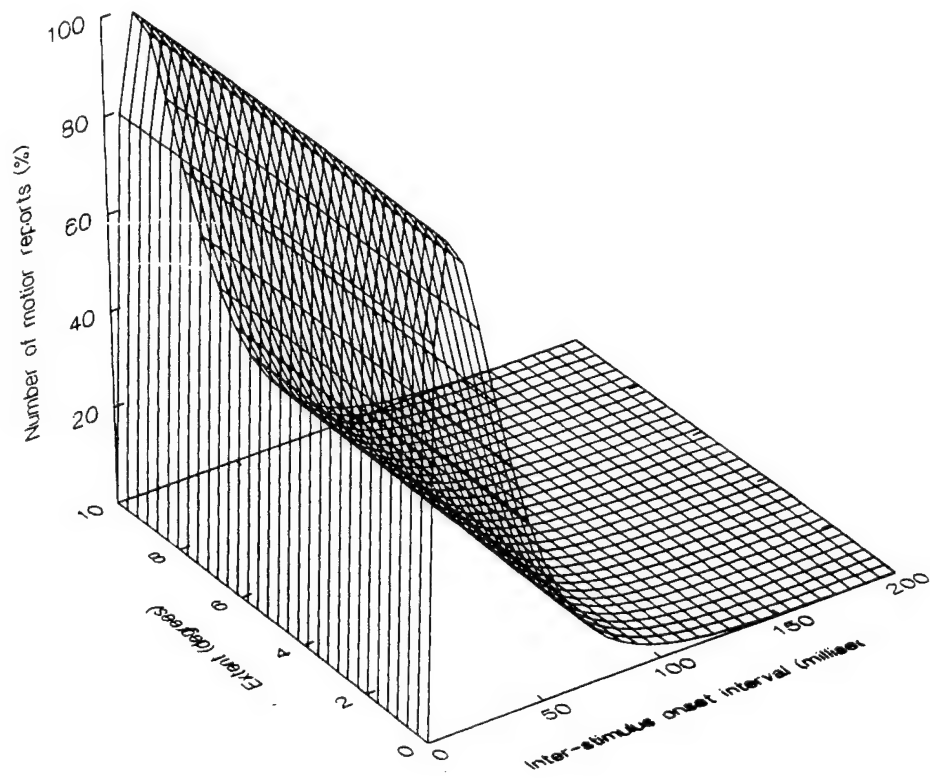


Figure 3.19: The representation in Figure 3.17 was expanded to include extent as a variable using data from Strybel [120] [123] at a constant duration of 50ms.



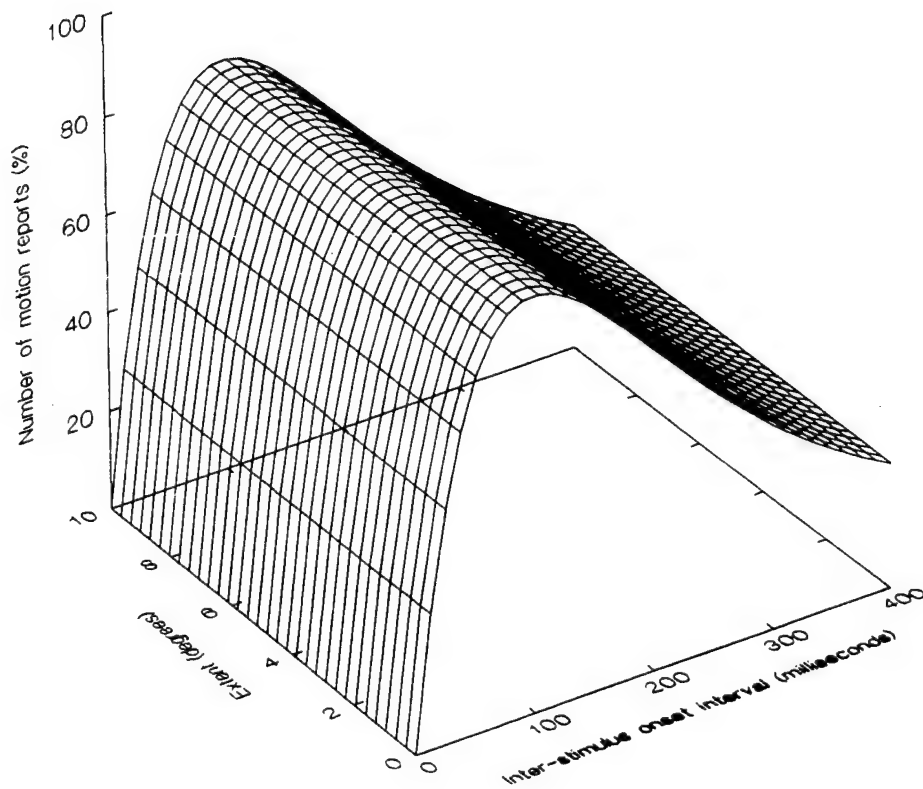


Figure 3.20: The representation in Figure 3.18 was expanded to include extent as a variable using data from Strybel [120] [123] at a constant duration of 50ms.

corresponds to approximately 7Hz. The gain of the temporal filter utilized by Watson has a maximum gain at approximately 7Hz as well as a steep slope from the higher frequencies to 7Hz and a much more gradual slope from the lower frequencies to 7Hz. A. J. van Doorn and Koenderink stated that the average delay in the temporal filters within the Reichardt detector model was 66ms [24] [23]. The delay of 66ms is associated with movement of approximately 7.5Hz by doubling and inverting the resultant delay. A.J. van Doorn and Koenderink also state that there is evidence that the range of delay and separation of the Reichardt detectors in the human visual system can be considered continuous [24].

### **Image complexity influence characterizations**

There are many other contributors to the perception of visual and auditory apparent motion than temporal and spatial extent that are relevant to experimental work described within this report. These other contributions must be controlled experimentally if the affect of ISI and spatial extent are to be manipulated as independent variables in an experimental setting. In general, these contributions fall into the category of visual and auditory complexity.

Some work had been done regarding visual image complexity perceptual effects. Cutting and Garvin utilized fractal curves [89] [68] to modulate perceived image complexity [20]. They found a correlation between the fractal dimension and several variables commonly used in the psychophysical literature, including symmetry, moments of spatial distribution, angular variance, and number of sides [20].

Raymond studied the interaction of target size and background pattern on visual motion perception. He found that target size increases as well as certain background contours can cause an under estimation of target velocity [99].

Braddick found that when using overlapping random-dot patterns separated by 15' of arc, motion perception could be elicited for SOAs less than 100ms. However, Braddick also found that isolated single spots could elicit motion perception with SOAs of up to several hundred milli- seconds [14].

Breitmeyer, May and Williams evaluated how spatial frequency and contrast affect apparent motion perception. They concluded that the perception of apparent motion decreased as a function of increased spatial frequency and increased as a function of increased contrast [15]. Finlay and Von-Grunau studied the breakdown effect of the perception of stroboscopic motion using 1-dot 2-frame stroboscopic stimuli, separated by 2 to 4 degrees, and switching positions between 0.75Hz and 6.0Hz. At the 3Hz rate, the perception of apparent motion would breakdown after viewing the stimulus for approximately 20 seconds [30]. They concluded that the size of the target was not a significant source of the breakdown, but spatial separation and temporal frequency did contribute to the presentation time necessary to cause breakdown of the apparent motion perception [30].

The shape of individual figures used in stroboscopic stimuli did not appear to have

a large bearing on the visual apparent motion perception. After reviewing many of the studies in this area, Kolers concluded that figural identity had a limited role in apparent motion perception [69]. Grossberg stated that more recent studies had confirmed that finding [45]. Grossberg cited work by Kolers and Pomerantz in 1971 in which they found that, when manipulating ISI and shape as independent variables and measuring the probability of reporting motion perception, that shape dissimilarity accounted for only 1% to 3% of the statistical variance [45].

Sekuler performed research regarding size effects within competing visual apparent motion stimuli. He found that size was a contributor to the perception of visual apparent motion and that the processing to extract size information must have been an early processing stage [105].

Kaufman and Williamson studied the ability to perceive visual acceleration using sinusoidal gratings and random dot patterns. They concluded that detection of changing direction was a higher level cognitive function than detection of motion itself [62].

Hubbard and Bharucha described an evaluation of five experiments to judge the apparent visual vanishing point of a target traveling in a two-dimensional space. They found significant deviations from actual and estimated vanishing points [57]. They also concluded that these deviations were caused by high level cognitive mechanisms [57].

### **Long-term temporal effects**

There is evidence within the literature of temporal effects within the visual and auditory motion perceptual systems that are much longer than the time required to detect motion. These temporal effects range from approximately 500ms to 25 seconds. These temporal effects are manifested as perceptual after-effects, adaptation, hysteresis, and perceptual recruitment effects.

There have been several studies of auditory motion after-effects from horizontally moving sound sources. Grantham concluded that the auditory motion after-effect stems from a generalized bias towards reporting opposite movement between probe and adapter and that a loss of sensitivity to motion occurred after prolonged exposure to moving sounds having similar spectral content [40]. The procedure utilized by Grantham involved a sequence of trials in which an adapter and probe were alternately presented. The adapter moved repeatedly through the subject's front hemisphere. It moved in a semicircular arc, with a constant angular velocity, during the sequence of trials. The probe was also presented in the front hemisphere but with randomly selected velocities through the sequence of trials. The subject was asked to determine the direction of the probe's movement after its presentation.

Grantham found in a later study that there was potential evidence of both a short-term adaptation effect and a long-term adaptation effect resulting from the presentation of a moving adapter prior on the minimum audible movement angle

(MAMA) [41]. In these studies, Grantham found that the MAMA was increased significantly by utilizing a moving probe preceding the auditory stimulus [41]. Grantham also found that the MAMA, with no probe present prior to the stimulus, obtained before a set of trials was significantly smaller than the MAMA obtained with no probe present prior to the stimulus, after a set of adapter probe trials. This indicated to Grantham that there was long-term and short-term motion detection adaptation occurring [41].

The perceptions of visual motion after-effects were studied by Hershenson and provide an interesting insight into the possible nature of tuned-detectors for both rotation and size-change. He concluded that both of these types of detectors were stimulated by rotating Archimedes' spirals [51].

Williams *et al* utilized random dot patterns, consisting of 512 dots over a 16° diameter circular display, and required subjects to report on the direction of motion of the dot patterns [142]. In these studies, the proportion of dots moving in a consistent direction was increased over time until the subject reported motion and then was decreased until the subject reported no motion. The viewing time between each perceptual switch was approximately 10 seconds and resulted in hysteresis in the proportion of dots required to elicit motion perception [142].

The breakdown of apparent motion perception after prolonged viewing of stroboscopic displays may be associated with fatigue of neurons involved in apparent motion perception [30]. Finlay and Von-Grunau concluded that spatial separation and temporal frequency of the stroboscopic display did contribute to the presentation time necessary to cause breakdown of the apparent motion perception [30].

Eggleston described how prior perceptions affect the perception of visual apparent motion. When an ambiguous matching stroboscopic stimulus was presented, the perceptual matching was affected by prior apparent motion perceptions as well as spatial-temporal aspects of the stimulus [27]. In addition, Eggleston developed a novel psychophysical paradigm called the Method of Interleaving Anchors that appeared to be useful in distinguishing prior correspondence effects from response bias in the form of habituation error [27]. The Method of Interleaving Anchors was described by Eggleston as including the presentation of anchor displays within a set of test displays in a random order. The Anchor displays differed from the test displays in that the Anchor displays provided a stable and known percept while the test displays provided ambiguous apparent motion correspondence solutions. By interleaving anchor displays with test displays, the influence of a known prior correspondence problem that was offered by the anchor displays could be determined [27].

Eggleston's research was followed by studies reported in the literature by Anstis. Anstis performed several studies which evaluated the ability of stroboscopic stimuli presented prior to a moving stimulus to affect the apparent motion perception of that stimulus [3]. Anstis termed the effect of the priming stimuli *visual inertia*, and found that *visual inertia* did exist in the visual apparent motion perception system [3]. Anstis used an ISOI of 166ms and priming dots which appeared only

once prior to the evaluation stimulus. In these studies, Anstis showed that priming does not require a long exposure to a moving stimulus to elicit *visual inertia* [3]. Anstis also found that increasing the spatial separation between the priming stimuli and the evaluation stimuli reduced the *visual inertia* and that non-moving priming stimuli did not create *visual inertia* [3].

Pinkus supported the results obtained by Eggleston and Anstis by finding evidence of visual motion priming in the humans apparent motion perception system [98]. Pinkus took the visual motion priming evaluation further by studying the additional spatial and temporal characteristics of the priming effect [98]. Pinkus stated that priming can occur over a range from 190ms to 770ms [98]. Pinkus also stated that visual motion priming can be modeled, within the context of the elaborated Reichardt motion detector, as a Gaussian bandpass filter,  $F_t$ , replacing the temporal integrator [98]. In addition, Pinkus placed the Gaussian bandpass filter,  $F_t$ , after the summation stage of the elaborated Reichardt detector model [98]. The shape of  $F_t$  was described by Pinkus as a Gaussian impulse response given by the equation below [98].

$$F_t(t) = t^s e^{-dt} \quad (3.21)$$

In this equation,  $s$  and  $d$  are constants which determine the steepness and decay of the filter and can be selected such that the delay of the filter is on the order of several hundred milliseconds [98]. Pinkus did not specify a single value for  $s$  or  $d$  in his research.

### 3.1.5 *Intra-modal and inter-modal models of visual and auditory apparent motion perception*

It is clear from the literature that a single, comprehensive model of the visual or auditory motion perception system has not been constructed and would be extremely complex if it existed. While several quantitative models of the visual motion processing system have been developed, they all differ in levels of abstraction from a neuron-based structure, and none of them capture the complexity of processing in total. In addition, quantitative models of the auditory motion perception system were not documented in the literature.

A new model of visual and auditory apparent motion perception was created within this report which was derived from the Reichardt visual motion detector model. This new model incorporates elements, which may exist as cortical processes within the human, representing mechanisms of visual and auditory apparent motion perception elicited from simple stroboscopic stimuli and thus, is called a *mechanism model*. The mechanism model is more speculative in the auditory modality than in the visual modality because it was derived from the Reichardt model, a model of visual motion detection.

The visual and auditory mechanism model was further reduced within this report,

using inter-stimulus interval and horizontal extent as stimulus characteristics, into a visual and auditory *process* model. The *process* model uses inter-stimulus interval and horizontal extent of simple auditory and visual stimuli as inputs. The mechanism model and the process model are described more fully in the following sections.

### Mechanism model

By considering the simplest visual stroboscopic stimulus, a one dot to one dot pattern, and two physical characteristics of that stimulus, spatial extent (separation), and temporal interval, structures underlying the model depicted as Figure 3.1 were specified to form a new model. This new model is called the mechanism model. The kernel of the mechanism model is the simple Reichardt visual motion detector, which is shown, in an elaborated form, in Figure 3.9. The least complex form of the Reichardt detector is shown in Figure 3.21. This detector uses two receptors, centered about the *position* of the detector, connected through a pure delay element, a multiplier element, and a summing element. The *position* of the detector, in this context, refers to the area in visual or auditory space in which the detector is sensitive.

The kernel itself is directional in that the vector formed by the center of the two receptors forms the response direction. The directional tuning of the motion detectors modeled in the literature were implemented in many forms, from spatial filters to discrete components, such as Reichardt detectors, having individual directional characteristics. The pooling of multiple detectors having varied directional tuning enables the detector to compute motion of stimuli moving in arbitrary directions. The mechanism model developed in this report is simplified with respect to direction-tuning in that it computes motion in only a single axis.

The effects of inter-stimulus interval on apparent motion perception can be modeled using the structure of this kernel. However, the effects of extent on apparent motion perception can not be completely modeled using this structure alone.

To model the effect of extent, additional detectors are spatially-combined in increasing separations around the position of the center detector. The spatial separation gives each detector a unique speed-tuning characteristic in conjunction with the fixed delay elements. Sereno utilized this speed-tuning approach in a neural-network-based model [106]. Sereno utilized speeds from  $4^{\circ}\text{sec}^{-1}$  to  $128^{\circ}\text{sec}^{-1}$  spaced at one-octave intervals with one-half response widths of 3 octaves [106]. Combined with the average delay of 66ms in each detector, the largest Reichardt detector would span a distance of approximately  $8.5^{\circ}$  and the smallest would span approximately  $0.26^{\circ}$ . In Sereno's model, there is detection outside of the band of speed-tuning but it rolls off on both the low and high sides [106].

The strength of the apparent motion sensation from each detector in the mechanism model is weighted based on receptor separation in a fashion to reduce the weight as the separation increases. This weighting is required in this model to approximate the functions obtained by Strybel [120] [123]. The mechanism model is

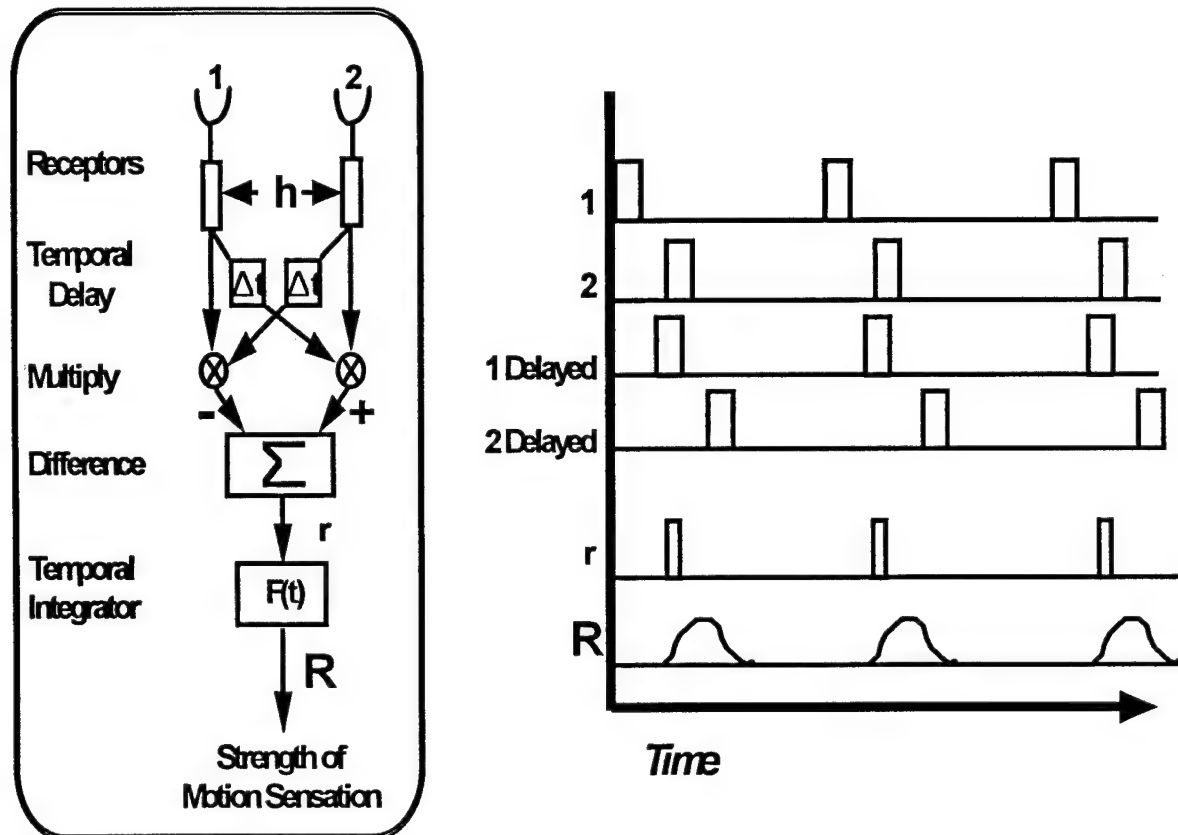


Figure 3.21: The kernel of the mechanism model, shown in the left portion of the figure, is a Reichardt visual motion detector. This detector uses two receptors, centered about position of the detector, connected through a pure delay element, a multiplier element, and a summing element.  $R$  is the strength of the motion sensation from the Reichardt detector kernel,  $\Delta t$  is the delay element, and  $h$  is the separation of the receptors. The right portion of this figure depicts an example of the timing of a stroboscopic stimulus arriving at receptor 1 and receptor 2 and the motion sensation strength,  $R$ , resulting from processing within the Reichardt detector.



shown in Figure 3.22.

In Figure 3.22, the  $\Delta t$  is equal *within* each Reichardt detector kernel and is equal *across* the Reichardt detector kernels. The separation between receptor 1 and receptor 2 is less than the separation between receptor 3 and receptor 4. Each kernels output is weighted based on the size of the separation and summed together resulting in an overall motion sensation strength across a specific spatial extent.

The plausibility that structures within the mechanism model exist within the human is supported by physiological structures that have been found in mammals. Cortical cells do exist within the cat and the monkey visual systems that exhibit direction selectivity and velocity selectivity [111] [76]. Velocity selectivity is achieved in the Reichardt kernel by the combination of the delay element and receptor distance. Direction selectivity is modeled by the single axis orientation of the Reichardt kernels. Direct physiologic evidence of the Reichardt structure in the human auditory system, which includes the sense organs as well as all of the intermediate and central processing supporting motion perception, was not found in the literature. However, auditory direction selectivity and motion sensitivity have been identified in owls and cats [110] [66] [65].

Physiologically-encoded topographical representations of auditory space have been identified in animals by several researchers [64] [70] [66] [65] [85]. There is also physiologically-encoded topographical representations of visual space within the macaque striate cortex [111].

The relationship of the mechanism model to models in the literature of visual motion perception is relatively close. In contrast, the appropriateness of applying the mechanism model as a model for auditory apparent motion perception is speculative. However, it is supported by the similarities between visual apparent motion perception and auditory apparent motion perception of simple stroboscopic stimuli, the fact that both modalities exhibit topographical mappings, and the physiologic level of processing in both modalities is cortical in nature.

The models depicted in Figures 3.21 and 3.22 are simplified relative to the actual processing that takes place in the cortex. The quantitative perceptual characteristics resulting from processing of inter-stimulus interval and horizontal extent within the mechanism model, along with data from the literature provide the foundation for building a model abstracted from the mechanism model. This model is formed within this report around the apparent motion perception processes involved with the inter-stimulus interval and extent of the stroboscopic stimuli. This model is called the process model and is developed within the following section.

### Process model

The cortical processing depicted in Figures 3.21 and 3.22 can be abstracted by recognizing that the effect of inter-stimulus interval on the apparent motion perception strength is separable from the effect of horizontal extent on the apparent



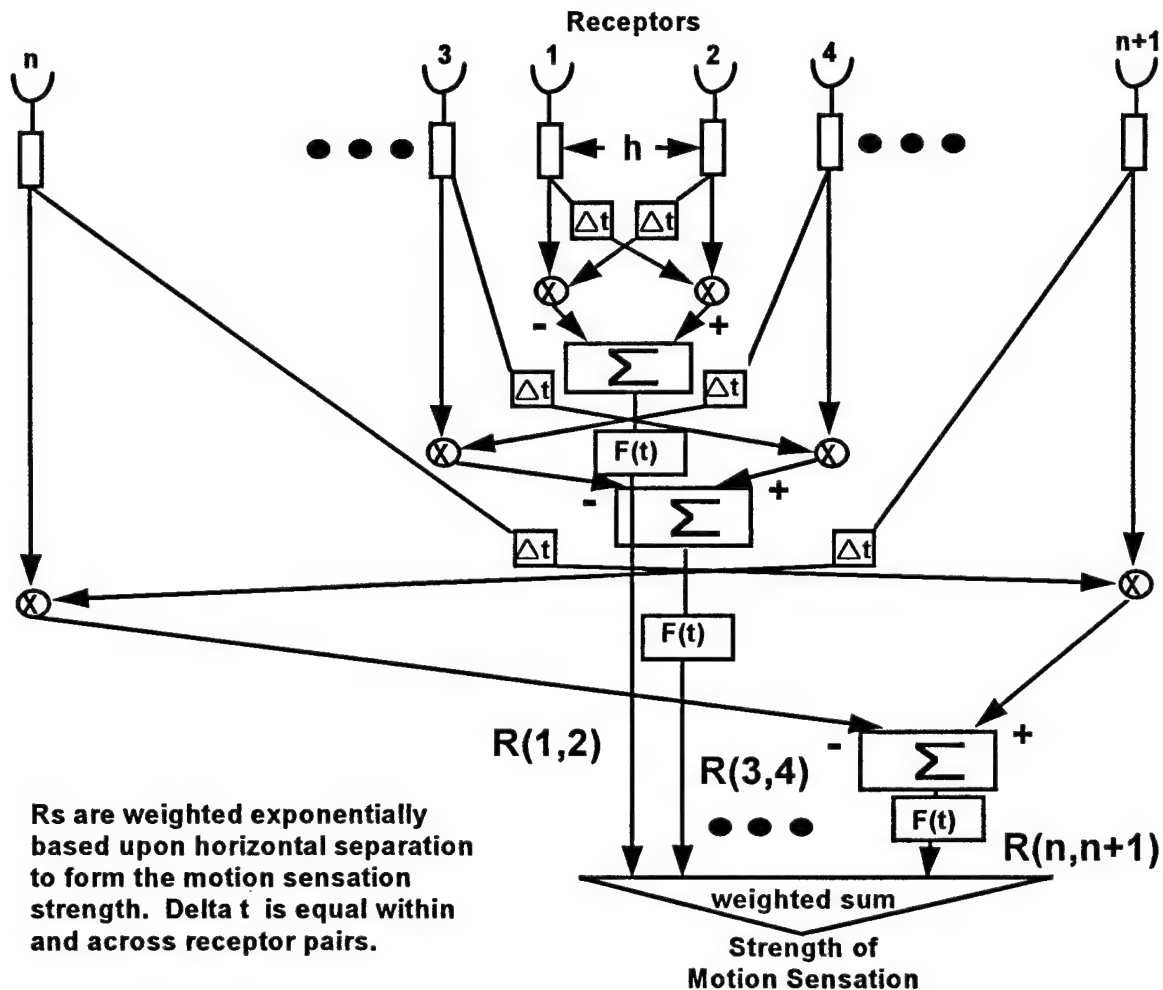


Figure 3.22: For motion in a single axis only, additional Reichardt motion detectors are arranged in increasing separations around the position of the center detector, decreasing in weight exponentially based on receptor separation, to account for spatial separation effects.  $R$  is the strength of the motion sensation from each of the Reichardt detector kernels. The  $\Delta t$  is equal within each Reichardt detector kernel and is equal across the Reichardt detector kernels. The separation between receptor 1 and receptor 2 is less than the separation between receptor 3 and receptor 4. Each kernels output is weighted based on the size of the separation and summed together resulting in an overall motion sensation strength across a specific extent. This is speculative as a model of auditory motion processing.

motion perception strength. This separateness is recognizable through analysis of Figure 3.22 as well as the data from Strybel [120] [123]. The abstraction of these models is consistent with the form of a process model of auditory and visual apparent motion perception that includes processing within both the intermediate and central levels. This model is shown in Figure 3.23.

The process model inputs the stroboscopic stimulus characteristics of inter-stimulus interval and horizontal extent and outputs measures of the characteristics, or qualities, of the apparent motion perception in each modality individually. A similar form exists for both the auditory and visual modalities. The specific elements in the model are the inter-stimulus interval processor, the horizontal extent processor, the multiplicative combination of the inter-stimulus interval processor output with the horizontal extent processor output into an internal representation of apparent motion perception strength, and a decision processor.

The inter-stimulus interval processor models the kernel Reichardt detector shown in Figure 3.21. The horizontal extent processor models the horizontal organization of the Reichardt detectors shown in Figure 3.22. The decision processor models the cognitive, or central, processing required to form a judgment of the apparent motion perception quality. Perceptual hysteresis, or perceptual capture, that is documented in the literature, is modeled by the temporal filter  $F_t$ .  $F_t$  replaces the temporal integrator which is a component of the elaborated Reichardt motion detector. The shape of  $F_t$  in this model is taken from Pinkus [98].

The velocity selectivity of motion detection described within the literature is inherent in the structure of the process model and the mechanism model. However, the range of direction selectivity and the various center locations of motion detectors within the human perceptual system can not be captured in either the mechanism model or the process model as described in Figures 3.22 or 3.23. However, direction selectivity and multiple central positions can be modeled with the mechanism and process models as a two-dimensional array of mechanism or process models. This two dimensional array would span the perceptual field of regard and at each center location multiple detectors would exist, each being *tuned* to a different direction.

The literature regarding visual motion perception models describes many methods for combining individual motion detectors like the mechanism and process model into unified models of motion perception across an entire field of regard. However, the number and spatial span of directionally-tuned detectors within the human visual system are not modeled consistently within the literature. As examples, Sereno modeled the direction selectivity as equal angular steps of 15°s with a one-half response of 60° [106]. Wang *et al*, in Durbins book, utilized 16 steps to cover the 360° range resulting in a 22.5° angular step size [25]. Watson utilized 10 steps to cover the 360° range resulting in a 36° angular step size [133].

Watson utilized the Watson-Ahumada detector model to form, what he termed, a vector motion sensor [133]. Watsons vector motion sensor was a combination of motion detectors in which the largest motion strength from any detector at a particular location determined the strength of motion at that location [133]. Watsons

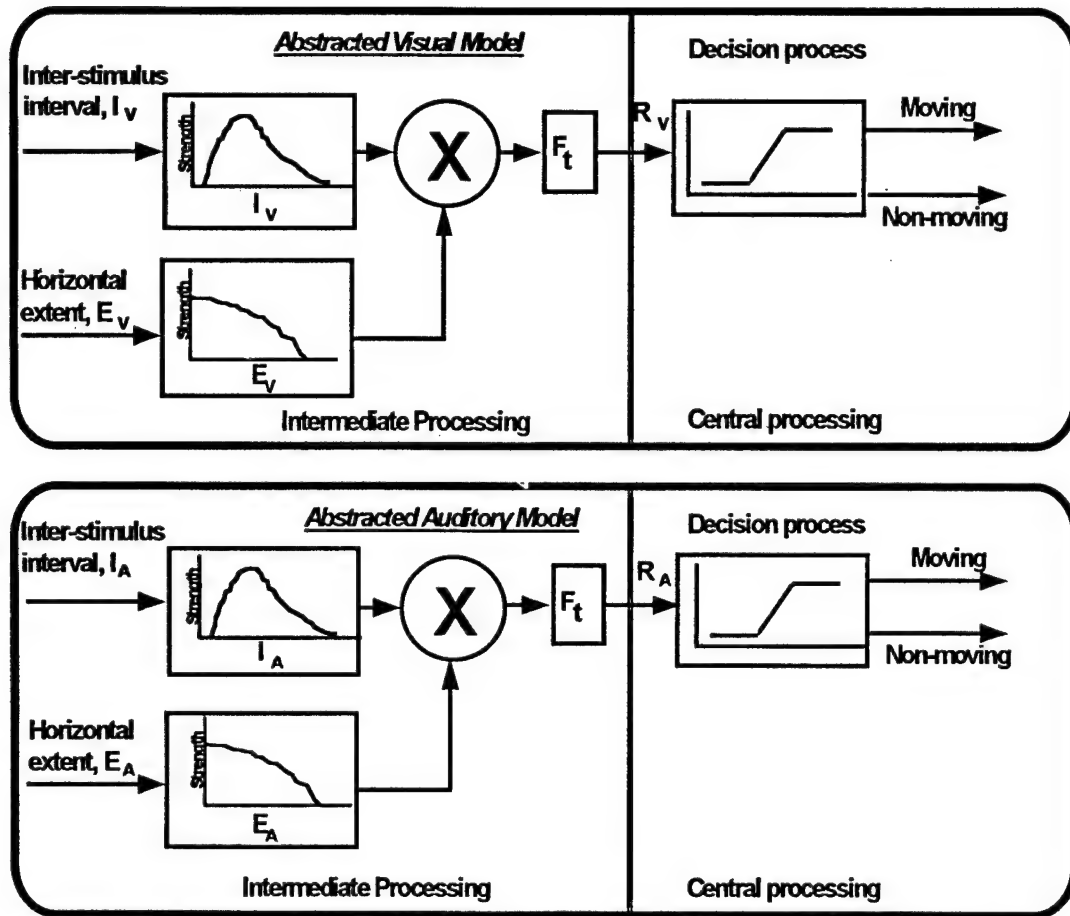


Figure 3.23: The cortical processing depicted in Figures 3.21 and 3.22 was abstracted by recognizing that the effect of inter-stimulus interval on the apparent motion perception strength is separable from the effect of horizontal extent on the apparent motion perception strength. This separateness is recognizable through analysis of Figure 3.22 and the data from Strybel [120] [123]. The abstraction of the mechanism model forms the process model of auditory and visual apparent motion perception. The inputs to the process model are the inter-stimulus interval and horizontal extent of the visual and auditory stroboscopic stimulus. The output of the process model is the characteristics, or qualities, of the apparent motion perception. The processor elements in the models are the inter-stimulus interval processor, the horizontal extent processor, the combination of the inter-stimulus interval processor output with the horizontal extent processor output into an internal representation of apparent motion perception strength, and a decision processor. The inter-stimulus interval processor models the kernel Reichardt detector model shown in Figure 3.21. The horizontal extent processor models the horizontal organization of the Reichardt detector models shown in Figure 3.22.  $F_t$  is a filter having a Gaussian impulse response modeled from the work of Pinkus [98].

model utilized a grid of vector motion sensors across the field of regard which did not interact with each other at the detector combination level [133]. Santen and Sperling also depicted visual motion perception using a grid of elaborated Reichardt motion detectors [129]. Santen and Sperling utilized a *voting rule* to combine the outputs of a grid of elaborated Reichardt detectors to form what they called an elaborated Reichardt model. Yuille, in contrast to Watson and Santen, modeled the combination of motion detector outputs using a two-dimensional grid of Gaussians centered on the central location of each motion detector [143].

The favorable comparison of attributes between the equations and literature-derived data, and between the mechanism model and process model supports the mechanism and process models as valid models of apparent motion perception as affected by  $I$  and  $E$  for simple stimuli in the visual and auditory modalities. In addition, the combination of these models into a grid representing a field of regard larger than a single detector is consistent with other models in the literature.

## 3.2 Inter-modal literature

The previous sections dealt with the visual and auditory apparent motion perception as independent systems. This section reviews pertinent literature regarding how visual and auditory motion perception may interact and influence one another.

### 3.2.1 Physiological basis for visual and auditory interactions

The research providing physiological information regarding the perception of combined auditory and visual motion involves inspecting cell responses to combinations of visual and auditory stimuli. This research appears to have been conducted with animals and tends to revolve around specific anatomical areas.

Animal studies performed by Meridith, Stein, Arigbede, Gordon, Chalupa, Dixon, and Rhoades all described cells within the superior colliculus of various animals, such as cats and hamsters, which responded to either visual, auditory, or multi-sensory stimulation [81], [113], [37], [21], [19]. The response to one modality could be greatly enhanced, (up to 326%), or reduced by the presence of stimuli in a separate modality [81]. This interaction in the animal central nervous system provides support for the possibility that similar cellular structures exist in humans and that interactions at the cellular level could form the basis of higher level processing at the perceptual and cognitive levels.

Several researchers found intra-sensory topographical encoding present in animals. Konishi and Takahashi [70] found that the Barn Owl maintains this mapping of auditory space. Knudsen found this same mapping in the Barn Owl [64]. Knudsen and Konishi found cortical cells in the Barn Owl that respond only to sounds originating within small elliptical areas in space and that these cells were arranged to

form a map of auditory space [66] [65]. Sovijarvi and Hyvarinen found cells in the cortex of the cat which respond to sounds in specific areas in space as well as to sounds that move in specific directions [110]. Palmer and King found topographical maps of auditory space within the guinea-pig superior colliculus [85]. Palmer and King also found that both monaural and binaural stimulus components were located in location sensitive cells of the guinea-pig superior colliculus and that the pinna appeared necessary for construction of a spatially-correct topographical map [85].

Both intermediate level and central level processing may be involved in the processing of combined visual and auditory stimuli. One possible mechanism of auditory-visual motion perception interaction is inter-cortical communication between the auditory and visual areas within the human cortex. Gilbert, Tso, and Wiesel reported that there was a significant amount of horizontal communication within the striate cortex of the monkey [76]. This result was obtained by using cross-correlation analysis between cortical areas [76]. However, the horizontal communication described by Gilbert *et al* appeared to be centralized to the visual cortex. No visual cortex to auditory cortex communication was described by Gilbert *et al*.

Stein, in a book edited by Vanegas, described the superior colliculus as having two distinct laminae, an area of superficial laminae and an area of deep laminae [130]. Stein stated that it is well known that the superficial laminae of the superior colliculus respond mainly to visual stimuli but that the deep laminae are multi-modal [130]. Stein also cited research stating that lesions in the deep laminae of the superior colliculus have multi-modal consequences and have produced cognitive effects such as a profound lack of attentive ability [130]. In addition, Stein stated that information processing in the superficial layer is sent to the visual cortex and that the visual representation in the superficial laminae is topographic and that the deep laminae of the superior colliculus were connected to both the auditory and visual sensory systems containing a topographical map of visual and auditory in approximate registration with each other [130]. Stein and Carlson both stated that the superior colliculus has striking responsiveness to moving stimuli [130] [18].

### 3.2.2 Visual and auditory perceptual interactions

Research concerning the perceptual interaction between the aural and visual modalities has been performed over many years [47], [77]. The research of inter-sensory processing of motion perception appears to be bounded by two distinctly different research methodologies. The first methodology involves the study of the physiological and neurological structure and processing within humans and animals in an attempt to build descriptions of perception based on deterministic understanding of existing processes and functions. The second methodology involves the study of psychological phenomenon, such as cause and effect relationships between stimuli and perceptions, with little regard to underlying neural or physiological basis.

In addition, several researchers have utilized an approach bringing together results from these different methodologies into signal processing and mathematical models of human systems [143], [1], [129], [116], [128], [53], [9], [29], [61], [72], [100], [52], [54], [8]. The results of developing and evaluating signal processing and mathematical models have provided direction to researchers in related technical disciplines. An example of this is Stockham's image processing work concerning the development of computation algorithms for photographic image enhancement that was inspired by physiological studies of the human visual system [117], [116]. Advances in mathematical signal processing concepts and algorithms have also impacted the ability to construct and evaluate models of the human visual system. An example of this is the work by Jasinchi involving space-time filtering [58].

London compiled an interesting account of inter-sensory research that had occurred in the Soviet Union prior to 1954 [75]. While London's account of the Soviet researcher's findings was sketchy at best, the diversity of visual-auditory interactions that he reported was noteworthy. London stated that the critical flicker frequency for green light was reduced by auditory stimulation while the critical flicker frequency of orange-red light was increased under auditory stimulation [75]. London also reported that auditory stimulation affects contrast sensitivity and brightness of visual after-images [75]. Also according to London, the Soviet researchers found that auditory sensitivity could be influenced by monochromatic room lighting, such that auditory sensitivity increases with green room illumination but decreased with red room illumination [75].

An interesting processing interaction study was performed by Pentti in 1955 using psychological methods [90]. Pentti published the results of a study in which a subject was seated inside a rotating black and white striped cylinder. The subject was stationary and the cylinder was rotated around the subject. An auditory stimulus was presented to the subject under moving and stationary conditions of the cylinder, with the subject being asked to localize and specify the location of the sound source. Pentti found that the ability of the subject to specify the location of the sound was significantly affected by the direction of the rotating cylinder [90]. Pentti also found the effect to be systematic, in that it appeared that the cylinder rotation shifted the auditory frame of reference in the direction of the rotation [90].

O'Leary and Rhodes investigated the cross-modal influences of perception of a multi-segmented stroboscopic visual display with perception of a multi-toned stroboscopic auditory display. The visual and auditory separations were not held constant for each subject but were tailored for each subject based on calibrations performed prior to the collection of data. O'Leary asked the subjects before each trial to respond to either the visual or the auditory display and recorded the SOA at the time of image segmentation. O'Leary found that the SOA that elicited segmentation of the visual display was altered by the presence of the auditory display and that the SOA eliciting segmentation in the auditory display was altered by the presence of the visual display [84]. Specifically, the SOA of visual segmentation was increased by 10ms, a 5% increase, with a two-object auditory display and the auditory segmentation was increased by 17ms, an 8% increase, with a two-object



visual display. Six of the eight subjects utilized in this study showed this effect.

The more recent cueing and attention studies tend to use reaction time as a measure of cueing or attention effects. These studies tend to indicate that inter-sensory interaction does occur within the human.

Miller studied reaction time effects in a divided attention paradigm and concluded that a co-activation model, not a separate decision model, accurately depicted the interaction [82]. Shelton and Seale concluded that vision improved the accuracy of auditory localization in the horizontal plane but did not affect accuracy in the vertical plane [108]. Stoffels and Van Der Molel investigated the effects of visual and auditory noise on visual choice reaction time. They found that by using a visual target, surrounded by visual noise, irrelevant auditory location cues could impair reaction time. In addition, they found that when auditory and visual noise was of the same general type, a cross-talk between the auditory and visual detection channels might occur producing perceptual conflict [118].

Perrott studied the choice-reaction time effects of auditory cueing on a location and identification task. Perrott found that reaction time could be reduced by utilizing the auditory cue, especially in elevated and rear quarter visual targets [91]. Gielen, Schmidt, and Heuvel also found reaction-time decreases under visual and auditory stimulation versus visual stimulation alone to be on the order of 20ms to 40ms [35]. Gielen also suggested that the decrease in reaction time could not be fully accounted for using a statistical facilitation model, which assumed that reaction time was based on the sensory modality that completed processing of a cueing stimulus first. Within this suggestion, Gielen stated that some inter-sensory processing must have occurred [35].

Shaw investigated the effects on detection decisions and information loss under conditions of simultaneous information presentation in the visual and auditory modalities [107]. Shaw found that separate decisions were formed about the presence of single tones when the tone of a specific frequency was defined as a single source of auditory information to the subjects [107]. When several tones were presented together, the decisions appeared to be made independently for each tone and summed at a higher decision level [107]. In addition, this independent decision model best accounted for information integration across modalities [107]. He also showed that no information was lost when multiple pitch channels were stimulated, losses did occur when multiple visual locations were involved in letter detection, and that losses of information did not occur if simple luminance detection was tasked [107].

Recent studies involving processing interaction between the visual and auditory modalities tend to measure functional performance effects due to sensory interaction.

Jones and Kabanoff evaluated the effect of eye movements, directed by visual targets, on auditory position reports. They found that eye movements tended to reinforce auditory position memory [59]. Welch, DuttonHurt, and Warren evaluated the interaction between aural and visual temporal rate perception and found that when both modalities were stimulated, audition provided a much stronger percept

than vision between 4Hz and 10Hz [134]. They also concluded that the naturalness of the modality for a given stimulus significantly affected the dominance of that modality in a visual and auditory presentation [134].

Wickens and Boles developed a theory for display presentation which involved multiple stimulus elements supporting the presentation of highly correlated information that must have been integrated into a single mental representation [139]. They stated that as the correlation of the information elements increased, the usefulness of separate presentations decreased [139]. Kobus and Lewandowski provided a practical example of such a correlated presentation in their simultaneous auditory and visual presentation of sonar information in a detection task [67]. They also concluded that simultaneous visual and auditory presentation aids task performance [67]. Barbour investigated the simultaneous presentation of visual and auditory signals buried in noise for a signal detection task [7]. The task was to identify what signal was present as the signal and noise were slowly separated. In two of the three signal patterns, the simultaneous auditory and visual presentation was superior [7]. For the third signal pattern, the auditory presentation alone was superior [7].

Auerbach and Sperling presented an evaluation that supported the hypothesis of the existence of a mental model of space that was common to both the auditory and visual modalities. They evaluated this hypothesis using a signal detection theory framework [4].

One of the major differences between visual and auditory apparent motion perception is in the ISOI values that elicit motion reports. Briggs and Perrott determined that motion in the auditory domain was produced by ISOI values ranging from -40ms to 40ms using a stimulus duration of 50ms. Neuhaus, in 1930, determined the ISIs needed to elicit motion reports, using stroboscopic visual stimuli of 40ms duration, to be between 60ms and 270ms [121].

One recent paper by Strybel *et al* illuminates the differences of the two modalities in their ability to elicit apparent motion perceptions. In one of the reported studies contained in Strybel's paper, it appeared that if there was any time between the cut-off of the initial auditory source and the turn-on of the second auditory source, it destroyed the appearance of continuous motion using a 50ms auditory stimulus duration [120]. The visual domain appeared to be much more tolerant of this interval, in that a time of 150ms could elapse before equivalent degradation would occur in the perception of continuous motion [120]. This is supported by results of Green in which gaps in noise as short as 3ms are easily detectable by subjects [42].

Warren, McCarthy, and Welch, and Stanislaw had developed or evaluated methodologies to be utilized in multi-sensory signal detection and interaction research. Warren provided support for the continued use of experimentally imposed discrepancy between modalities to aid in the study of sensory interaction [132]. Stanislaw presented a methodology that aided the distinction between the effects of divided attention and the effects of stimulation of one modality on the perception by other modalities. Stanislaw stated that they could be made distinct by presenting



supra-threshold stimuli to each modality on every trial, and by administering concurrent tasks in conditions involving divided attention [112].

### 3.2.3 A process model of auditory influence over visual motion perception

A model of auditory influence over visual motion perception must incorporate a detection mechanism, a perceptual/cognitive mechanism, and an interaction or influence mechanism. The literature does not conclusively drive an influence mechanism towards a specific form. However, constructing a model of a hypothesized influence mechanism contributes to the formulation of research hypotheses. To this end, a construct of the auditory influence over visual motion perception, which is a modification of Figure 3.1, is proposed and is shown as Figure 3.24.

The construct depicts characteristics of the visual and auditory stimuli, represented by  $I$  and  $E$  from *Korte's laws*, entering the sense receptors, either the eyes or the ears, on the left side of Figure 3.24, and perceptual characteristics reportable by the human observer, represented by  $C$ , exiting the construct on the right side of Figure 3.24.  $I$ ,  $E$ , and  $C$  may be measured through the use of instrumentation or dialog with the human observer. This form is similar to the intra-modal construct depicted in Figure 3.1.

Within the construct shown in Figure 3.24, two levels of processing are depicted, those levels being motion detection and motion perception. The variable depicted between these processing levels is the strength, or robustness, of the motion sensation. The strength of the motion sensation is represented by  $R$ .  $R$  can not be measured directly in the human.  $R$  is correlated with the characteristics of the stimulus and contributes to the perceptual and cognitive processes in the human necessary to respond to the stimulus. The subscripts  $a$  and  $v$  are used as subscripts on the variables  $C$  and  $R$  to represent the auditory and visual modalities respectively. The interaction, or influence, that may occur between the visual and auditory modalities is depicted in the center of Figure 3.24. The influence is shown as receiving input from the visual and auditory motion sensation strength and outputting a modified visual strength of motion sensation represented by  $R'_v$ . The definitions of the variables in the construct are reflected in Figure 3.24.

The construct depicted as Figure 3.24 can be built upon by inter-linking the process model depicted in Figure 3.23 into a speculative connection between the auditory and visual motion perceptual systems supporting the development of hypotheses regarding the effect of moving auditory stimuli on visual apparent motion perception. The inter-model connection hypothesizes an effect at the intermediate level in this figure consisting of a multiplication element. The intermediate level connection speculates that the Reichardt kernel is not directly affected by moving auditory stimuli. In this manner, the central processing would be affected by the increase in motion sensation created by the multiplication element. This speculative inter-modal model builds upon the process model structures is shown in Figure 3.25.

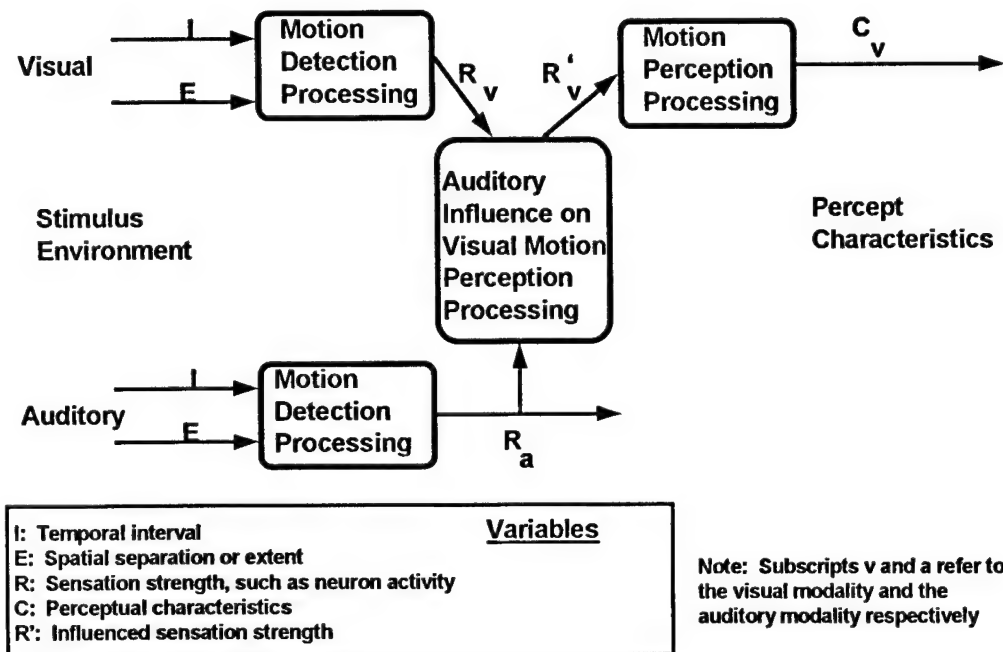


Figure 3.24: A construct of the auditory influence over visual motion perception was developed which is a modification of Figure 3.1. The influence is portrayed as receiving input from the visual and auditory motion sensation strength. The influence is portrayed outputting a modified visual strength of motion sensation represented by  $R'_v$ .

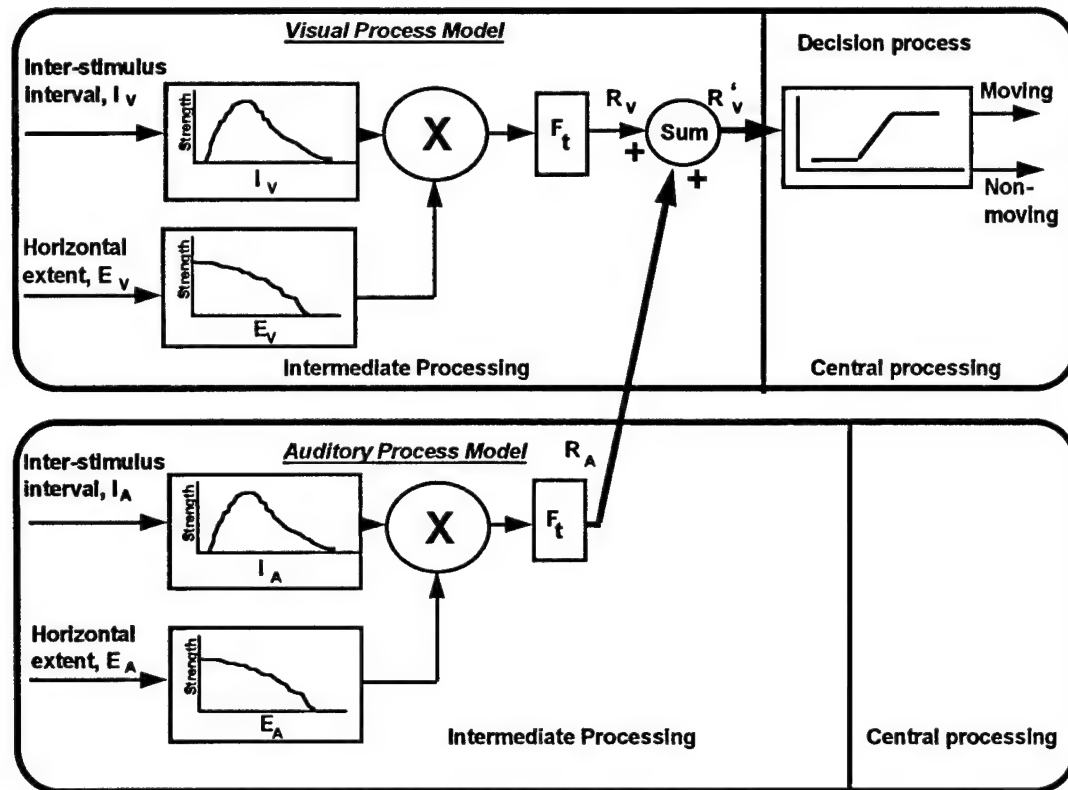


Figure 3.25: The construct depicted as Figure 3.24 can be built upon by inter-linking the process model depicted in Figure 3.23 into a speculative connection between the auditory and visual motion perceptual systems. The inter-model connection provides an influence pathway at the intermediate level in this figure consisting of a multiplication element. The intermediate level connection speculates that the Reichardt kernel is not directly affected by moving auditory stimuli. In this manner, the central processing would be affected by the increase in motion sensation created by the multiplication element.

The inter-modal process model, in conjunction with the mechanism model, provides a structure which supports identification of hypotheses within this report regarding the influence of stroboscopic auditory stimuli on visual apparent motion perception.

### 3.3 Summary of the literature review

There is by far much more literature regarding the *intra*-modal aspects of apparent motion perception than *inter*-modal aspects of apparent motion perception.

There is physiological evidence of visual-auditory interaction. The locations of the interacting information flow are not well defined. There appears to be an intermediate-level information channel within the superior colliculus supporting interactions between auditory and visual motion perception. There is also evidence of widespread horizontal communications within the cortex but no *inter*-cortical communications were described.

There is also psychological evidence of visual-auditory interaction. Some of these interactions, such as those classified as attention-based, appear to produce somewhat intuitive effects, while other interactions are not intuitive at all, such as Pentti's description of an apparent systematic reduction of auditory localization ability during and after the perception of visual motion.

Human motion perception, in both auditory and visual modalities, appears to be performed at several neurological levels and is both intermediate and central in nature. This is evidenced from the physiological studies as well as modeling efforts from both psychological and physiological perspectives. The majority of literature describing visual-auditory perceptual interactions could be described as processing interactions at the intermediate level within the superior colliculus. As an example of this, performance metrics based on reaction time or attention could be affected by processing within the deep laminae of the superior colliculus and also be multi-model in nature. In contrast to this, some inter-modal effects described within the literature, such as the visual motion interaction with auditory localization, do not appear to be examples of superior colliculus processing but do appear to be examples of central-level processing interactions.

The models developed within the literature to describe human motion perception ranged from models developed from psychological approaches to models developed from signal processing or mathematical approaches. Most of these models vary greatly in their outward appearance but maintain an underlying similarity in the functional description of several aspects of motion perception. A striking example of this is the Yuille-Grywacz computational theory of coherent visual motion and Petersik's description of the two-process distinction in apparent motion. Both of these models, the former being quantitative in nature and the latter being qualitative in nature, depict the perception of visual apparent motion as the combination and interaction of two separate processes.

No proposed processing models of the interaction between visual and auditory motion perception were found in the literature. All of the quantitative models of motion perception in the literature are of the visual system. No motion perception models were found for the auditory system.

Two new models were developed and described within this literature review, the mechanism model and the process model. The mechanism model was derived using the Reichardt motion detector as a kernel and as such, is tightly supported by visual motion perception literature but is highly speculative as a model of the auditory motion perception system. Also within this literature review, the mechanism model was reduced into an *intra*-modal process model of visual and auditory motion perception and was then modified to become an *inter*-modal model.

The *inter*-modal process model utilizes inter-stimulus interval and spatial extent as inputs to an integrated auditory and visual motion perception process. The *inter*-modal process model, as a three-stage processor model, supports the identification and assessment of hypotheses within this report regarding the influence of stroboscopic auditory stimuli on visual apparent motion perception.

## Chapter 4

# Apparatus and methods description

During the course of this empirical work, two different subject stations were utilized. The two stations, the Speaker/LED station and the Localizer/CRT station, were fabricated specifically to support the research described within this report. The Speaker/LED station was used in Experiment one and the Localizer/CRT station was used in Experiment two through Experiment eight. The Speaker/LED and Localizer/CRT stations are described in detail in the following sections.

Both stations were tied to an AT-class personal computer that served as the experimental controller and provided real-time datum collection and recording functions. Data analyses were performed off-line from the experimental facility using a VAX-785 and the SAS statistical software package as well as an 80486-class computer using the SYSTAT statistical-analysis software. Signal processing analyses were also performed on the 80486-class computer using the MATLAB analysis software.

### 4.1 Speaker/LED stroboscopic subject station

The Speaker/LED station consisted of a AT-class microprocessor controller with hard-drive and floppy drives, a light-emitting diode (LED) array configured in the shape of an "L", two small speakers, a subject response box with buttons, an audio signal source ( B&K audio noise generator, Type 1405), and an electronics interface box that switched the audio signals to the speakers and the power sources for the LED array. A LED is a semiconductor diode that emits light when a voltage is applied. A diagram of the overall station layout is contained in Figure 4.2. A diagram of the speakers and LED array is shown in Figure 4.1. Commands received from the digital output card resident in the microprocessor controlled the LED array and speakers through the electronics interface box. The electronics interface box also buffered the subject response buttons and fed them to the controller's digital input

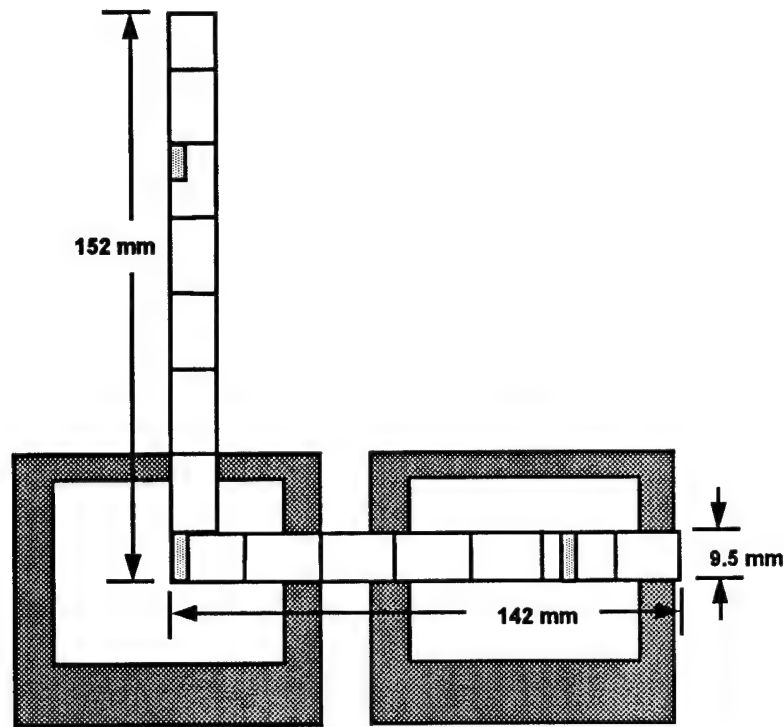


Figure 4.1: The LED array consisted of 7 quad LEDs along a horizontal row and 7 quad LEDs along a vertical row. This arrangement allowed 28 independent positions to be activated in the horizontal row and 14 independent positions to be activated in the vertical row. The LED array was rigidly attached to the two 4-inch speakers.

card.

The physical spatial characteristics of the individual speakers and lights were not alterable during a single trial. The LED array consisted of 7 quad LEDs along a horizontal row and 7 quad LEDs along a vertical row. On the vertical row, the LEDs were orientated vertically as opposed to the horizontal row, in which the LEDs were orientated horizontally. This configuration allowed 28 independent positions to be activated in the horizontal row and 14 independent positions to be activated in the vertical row. The speakers were 4 inch circular speakers [125].

Each speaker was mounted in a separate plastic case and the cases were rigidly positioned with respect to each other and the LED array. The entire speaker and LED assembly were rigidly mounted to the back of a desk. The subjects were seated at the desk in front of the speakers and LED bar. The subjects were positioned individually with a chin rest to maintain a single spatial relationship with the speakers and LED array. Sound absorbing material was positioned in back of the speakers and LEDs as well as along each side of the subject spanning the area from the speakers to the subject. The room was darkened during the collection of data.

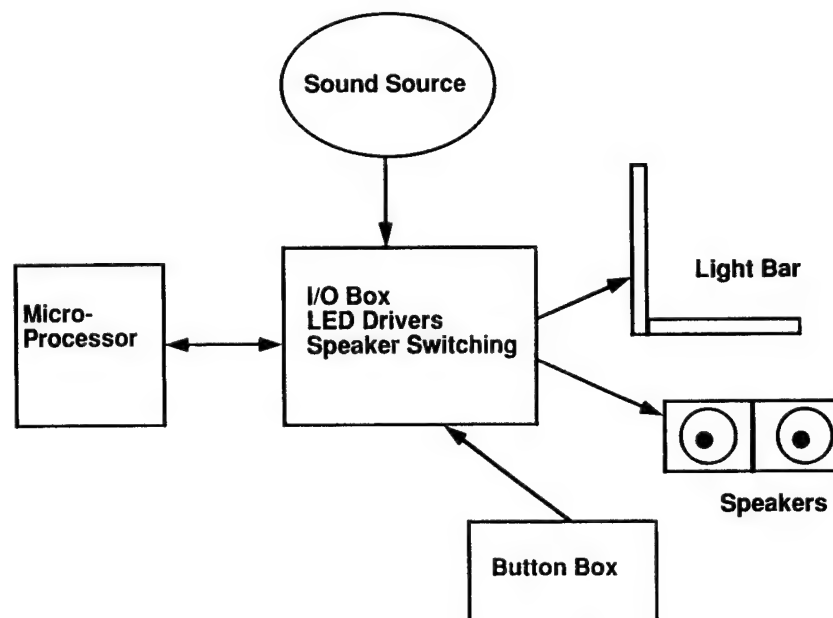


Figure 4.2: A parallel digital output from the micro-processor controller provided commands to electronic switches within the I/O box that gated the sound source as well as the LEDs.



## 4.2 Localizer/CRT subject station

The Localizer/CRT subject station provided the controllability and functionality required to produce stroboscopic auditory and visual stimuli, supported subject interaction with the stimulus, and recorded subject responses. The general layout followed the form depicted in Figure 1.2. The Localizer/CRT station is shown in diagram form in Figure 4.3.

### 4.2.1 Hardware components and characteristics

The display devices within the Localizer/CRT station were a high-resolution, high-speed, raster-scanned Tektronix CRT monitor model SGS 625, an audio signal source (B&K audio noise generator, Type 1405), an auditory localizer, and a set of high performance STAX headphones, model SRD-X. The CRT, auditory localizer, and headphone combination provided dynamic visual and auditory images under computer control. The auditory image was created by noise and signal generators under control of the microprocessor. The noise and signal generator fed the auditory localizer which created an audio stereo pair. The stereo pair was then portrayed to the subject through the STAX headphones. The auditory display was a virtual auditory display.

The auditory localizer integrated into the Localizer/CRT station was a high-speed digital signal processing device built around the Texas Instruments signal processing chip [80]. The localizer input was a monaural signal and produced a stereo pair as output simulating the audio input to the left and right ear canals of a monaural sound source emanating within an anechoic chamber. The output was produced through a bank of digital filters corresponding to a specific azimuth angle relative to the listener's head. A block diagram of the internal structure of the localizer is shown in Figure 4.4. A more detailed description of this auditory localizer is contained in the AFIT thesis by McKinley [80]. There were two potential auditory localizer systems available for use. The second localizer, The Convolotron (fabricated by Crystal River Engineering) was not chosen for this application. The major reason for this decision was that the Convolotron interpolated filter characteristics (modeling individualized head-related transfer functions) in real-time in an attempt to smooth the resultant auditory motion. In this research, the smoothing would have reduced the ability to quickly and precisely move the position of the virtual auditory source.

The head tracker utilized in the CRT/Localizer research station was a Bird system, built by Ascension, Inc. The Bird used magnetic transducers and provided position and attitude measurements in 6 degrees-of-freedom. The static positional accuracy was approximately 0.3 cm and the static angular accuracy was approximately 0.5 degrees. The Bird operated at 120 Hz with position and attitude data being supplied to the Z-248 under software control. While other magnetic-based head tracker systems were available, such as the Polhemus 3-Space system, the Bird system was used because of the relatively high operating rate of the system and its

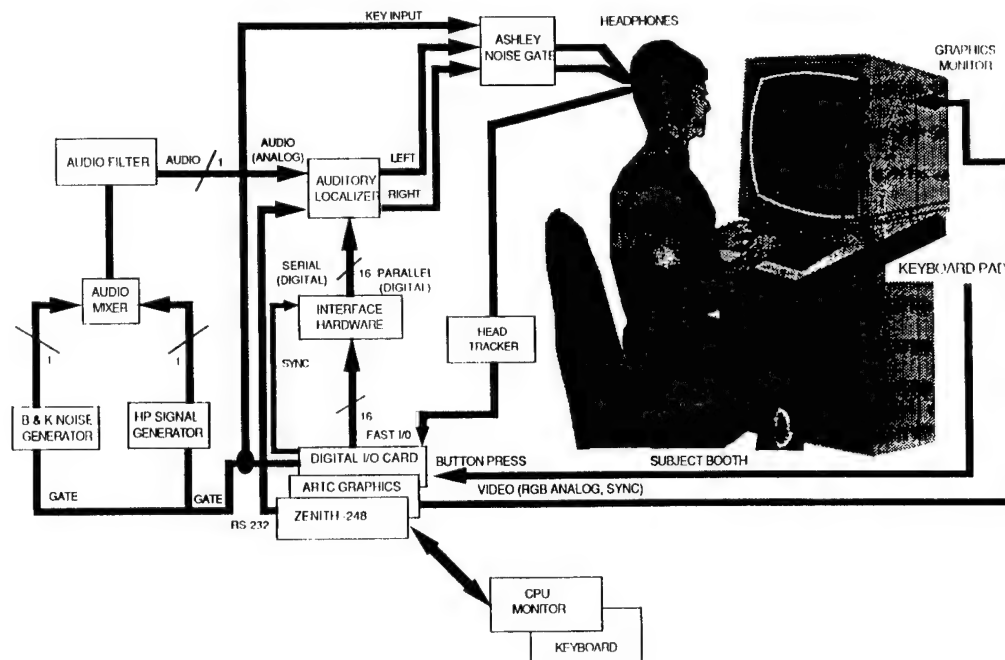


Figure 4.3: The micro-processor controller coordinated the generation of graphics to the CRT and positioning of the simulated auditory source.

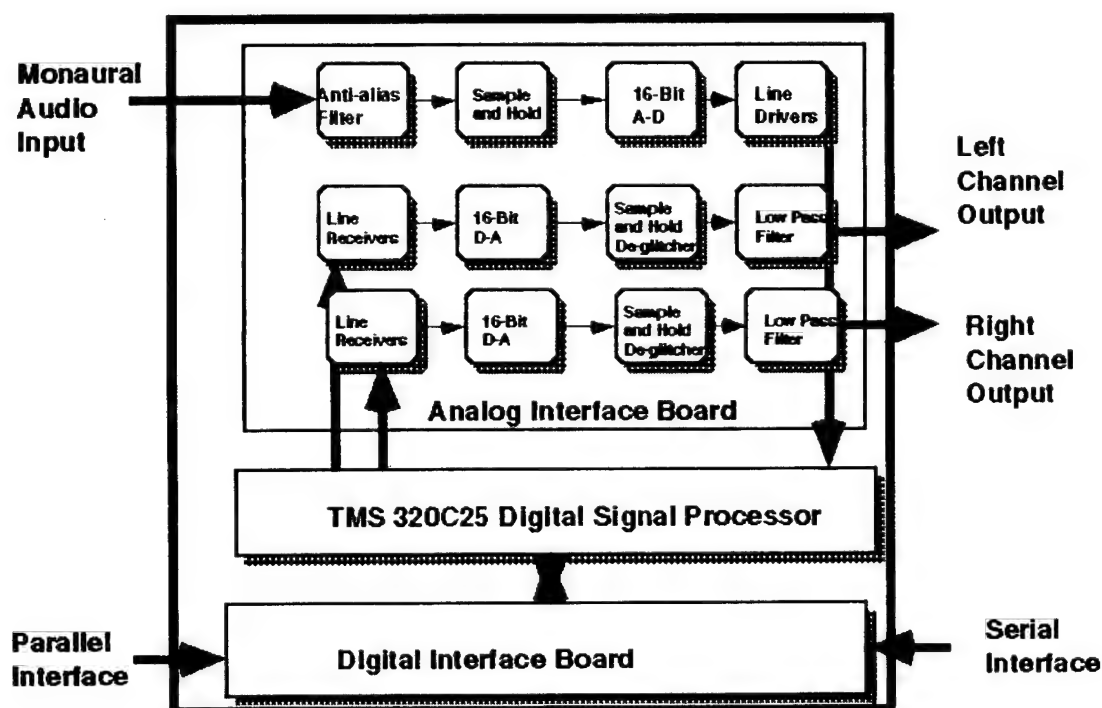


Figure 4.4: The auditory localizer integrated into the Localizer/CRT station was a high-speed digital signal processing device built around the Texas Instruments signal processing chip [80]. The serial interface was used for initialization and the parallel interface supported real-time manipulation of the simulated auditory source azimuth angle.

relative non-sensitivity to metal objects close to the magnetic transducer.

The microprocessor controller was a Zenith Z-248 micro-computer which controlled the generation of visuals by down loading graphics instructions to an internal graphics card that incorporated an ARTC graphics chipset. The visual display was of conventional construction and produced visual signals at temporal rates and spatial resolutions adequate to stimulate long-range visual apparent motion perception. Specifically, the raster-based visual display updated at 120Hz with 1024 separately addressable pixels over 36° when viewed from a distance of approximately 62 cm. (24.5 inches). The ARTC graphics chip set produced R-G-B signals which drove the Tektronix monitor.

The Z-248 also gated the B&K audio noise generator, Type 1405, and a HP function generator, type 8116A. The noise generator was passed to the auditory localizer through an Ashley 4 channel noise gate, type SG-35. The rise and fall times of the input to the localizer could be manipulated by the microprocessor through the noise gate. The localizer filtered the incoming audio signal in two parallel circuits, producing a stereo pair in real-time, and drove a noise gate that, in turn, drove head phones. The interface from the Z-248 to the auditory localizer was a 16 bit parallel cable from which the localizer sampled the sound location at a 1000Hz rate.

#### 4.2.2 Visual and auditory display synchrony

The localizer and monitor were synchronized through the 120Hz synchronization signal originating in the ARTC graphics card. This constrained the facility to a minimum cycle time of 8.3ms. It was important to maintain the temporal synchrony of the appearance of a visual display on the CRT monitor and the appearance of a localized auditory signal on the headphones. This required that the Z-248 drive the internal graphics card and the 16-bit parallel cable to the localizer consistently such that an approximately constant time differential could be maintained between the appearance of the visual stimuli and localized auditory stimuli. This was accomplished by performing datum collection functions within random access memory, not performing disk access during stimulus generation, and maintaining a strict order in the software structure regarding the display of the visual and auditory stimulus frames.

As an example of the software structure, the following order was used to illuminate a single dot on the monitor spatially and temporally linked with a localized auditory presentation. The software ordering used would have been as follows:

- 1 The software would initially blank the screen and then gate the sound off.
- 2 The software would pre-compute the location of the dot and corresponding azimuth angle command for the auditory localizer and save this into memory.
- 3 The software would sample the horizontal line number and wait for a line

number greater than the bottom of the visible raster. This step began the timing loop. The next items occurred before the next raster sweep began.

The software would draw the dot into the display memory. (The dot would not appear on the screen until the next raster sweep.)

The software positioned the localizer to the new azimuth angle by writing the azimuth angle to the parallel port.

The software would gate the sound source on. Once gated, the monaural audio signal was applied to the localizer and appeared at the headphones within 1ms.

- 4 The next raster sweep would begin. The dot would appear on the screen after the beginning of the raster sweep, depending on where it was placed on the screen.
- 5 The software would wait the pre-computed duration period. The duration period was an integer number of 8.3ms raster periods.
- 6 The software would blank the screen and then gate the sound off.

Using this software order, the theoretical maximum difference between the sound appearing at the headphones and the dot appearing on the screen could be no more than 7.3 ms assuming the software did not take any time to execute, the localizer was at the very beginning of its 1000Hz cycle, and the dot was placed at the very bottom of the CRT monitor. However, due to these extremely unrealistic assumptions, delays were always shorter than the theoretical maximum delay.

More typical numbers for this delay were obtained by measuring the delay within the experimental set-up. The delay was measured by applying a photo resistor to the monitor and driving a two-channel oscilloscope with the output of the photo-resistor and the audio signal going to the headphones on separate channels. This measurement was accomplished with a rectangular dot appearing at the center of the screen. The delay ranged from approximately 1ms to 2ms. The audio signal from the auditory localizer appeared in the headphones ahead of the appearance on the monitor of the visual dot. The duration of the delay was dependent upon the location of the dot on the screen as well as the size of the dot.

#### 4.2.3 Physical layout and controls

A single human subject was used in the facility at any one time. The subject viewed the monitor and listened to the headphones in a booth of inside dimension approximately 1 meter wide by 1 meter long by 1.8 meter tall. The subject was seated in a chair within the booth during experimental trials. The inside of the booth was covered in sound-absorbing material and was reasonably light-tight. The door on the booth was held closed by small magnets located on the door edges. Ventilation was provided in the booth by a fan located above the booth and an inlet

vent located in the lower rear of the booth. The subject responded to stimuli presented in the booth by manipulating control mounted in a small box that was connected via a cable to the Z - 248. The control box contained a joystick, buttons, and a slide control. A small incandescent lamp was located in the booth to aid entrance to and exit from the booth, with a light switch being accessible to the subject. During experimentation, the lights outside the booth were dimmed.

The joystick was manufactured by CTI Electronics Corporation, model number M1500ES, and supported 2 axis movement. The joystick utilized induction coils to sense the position of the stick resulting in a design that did not include any moving electronic components. The stick was approximately 76mm in length set in a base that was approximately 133mm by 89mm. The height of the base was approximately 51mm. The base could be held in the palm of one hand and the stick could be moved with the other hand. The stick force was constant with respect to displacement with no deadband. The stick was adjusted to provide a displacement force of approximately 1.5 ounces in both axes. The output of the stick, for each axis, was an analog voltage proportional to the stick displacement in that axis and ranged from 75% to 25% of the supply voltage. Maximum displacement of the stick was approximately  $\pm 27^\circ$  in both axes.

## Chapter 5

# Spatial correspondence influences

### 5.1 Experiment one: Effects of moving auditory stimuli on horizontal and vertical correspondence thresholds within a visual apparent motion display

The literature review reveals that the quality and type of motion perception resulting from stroboscopic visual and auditory stimuli can be affected by the temporal and spatial characteristics of the stimulus itself. The literature review also provides ample evidence in the potential of *inter-modal* influence. It is therefore, not too large a leap to hypothesize that motion perception resulting from one modality may influence the motion perception of another modality.

This experiment investigates the potential influence of stroboscopic auditory stimuli on visual stimuli which can be perceptually organized based on the visual stimuli's spatial characteristics. It is hypothesized in this experiment that the spatially-driven perceptual organization of stroboscopic visual stimuli can be influenced by the presence of contemporaneous stroboscopic auditory stimuli.

The visual stimuli utilized in this experiment was a 2-frame 1-dot/2-dot stroboscopic stimuli described by Kolars in his discussion of display attraction [69]. The stroboscopic visual display is shown in Figure 5.1. This figure depicts two frames of visual stimuli which alternate. The angular difference between the single dot in frame one and the lower dot in frame two is defined as the horizontal extent. The angular difference between the single dot in frame one and the upper dot in frame two is defined as the vertical extent. The horizontal extent is defined for both the visual and auditory stimuli. However, the vertical extent is not defined for the auditory display.

Kohlers described the perceptual organization elicited from the visual stimuli as being a function of a single spatial characteristic, that characteristic being the ratio of the horizontal extent to the vertical extent [69]. In this way, the 2-frame alternating stimuli could elicit perceptions of vertical motion or horizontal

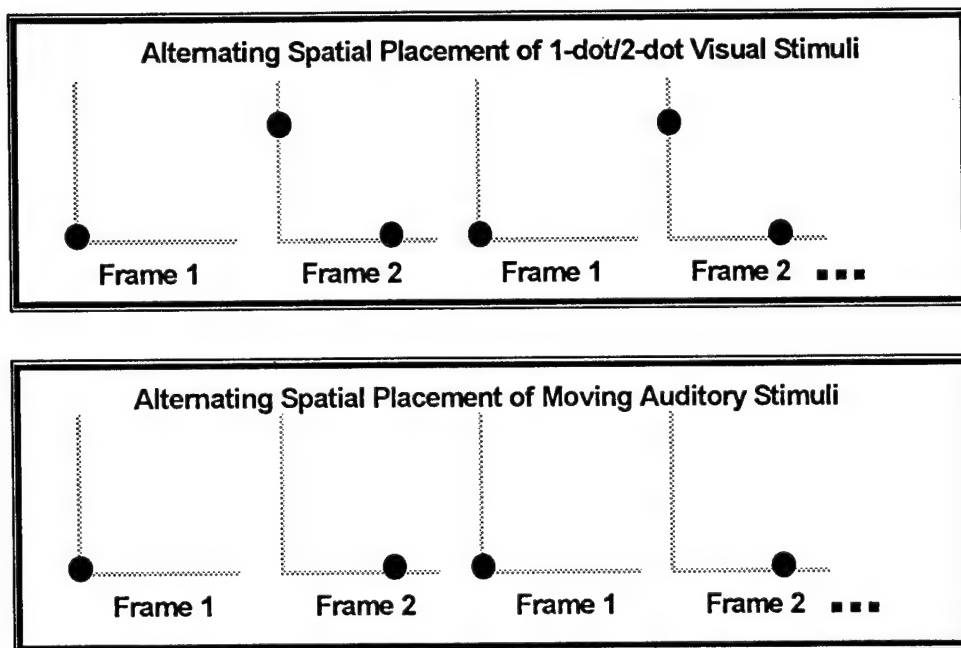


Figure 5.1: The visual stimulus used in this experiment was alternating one-dot and two-dot frames. The distance between the dot in the one-dot frame and the lower dot in the two-dot frame was systematically modified during the experiment. The ascending trials began with a small horizontal extent to vertical extent ratio yielding an initial report of *horizontal* motion. The descending trials began with large horizontal extent to vertical extent ratio yielding an initial report of *vertical* motion. The moving auditory stimulus was linked to the single dot in the one-dot frame and the lower dot in the two-dot frame. The stationary auditory stimulus remained at the position the single dot in the one-dot stimulus during the trial.



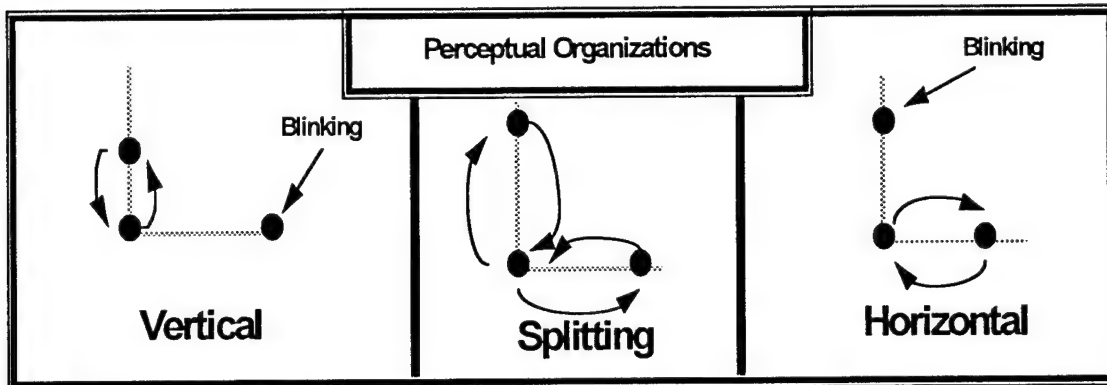


Figure 5.2: Based on the ratio of the horizontal extent to the vertical extent, the visual stimuli could be perceptually organized in three ways. For large horizontal extent to vertical extent ratios, the visual stimuli would most likely elicit vertical motion. For small horizontal extent to vertical extent ratios, the visual stimuli would most likely elicit horizontal motion. A third organization may be elicited if the horizontal extent to vertical extent ratio is close to unity. The third potential perceptual organization is characterized by alternating splitting and rejoining of a single dot into a vertical and horizontal dot.

motion [69]. The elicited perceptual organization would be driven by the strength of the vertical motion percept relative to the strength of the horizontal motion percept. In the vertical motion organization, a single dot would appear to move vertically with a second dot blinking to the right of the moving dot. In the horizontal motion organization, a single dot appeared to move horizontally with a second dot blinking above the moving dot. If the horizontal extent was equal to the vertical extent, a third organization might have been elicited. The third organization would appear as one dot splitting into both a horizontal dot and a vertical dot simultaneously. These three perceptual organizations are shown in Figure 5.2.

The hypothesis of this experiment, that the spatially-driven perceptual organization of stroboscopic visual stimuli can be influenced by the presence of contemporaneous stroboscopic auditory stimuli, poses the possibility that Kohlers' construction of perceptual organization strength could be affected by the presence of horizontally moving auditory stimuli. Specifically, it is hypothesized that the horizontally-moving auditory stimuli would increase the strength of the horizontal motion percept and that the ratio of horizontal extent to vertical extent that could

support the horizontal perceptual organization could be increased. Put another way, it is hypothesized that by spatially and temporally linking the auditory stroboscopic display to the horizontal orientation of the visual stimuli, the perception of horizontal motion will dominate the perception of vertical motion due to an increase in the strength of the horizontal motion percept and that the increased horizontally-orientated strength would result in increased reports of horizontal organization.

### 5.1.1 Aims

One objective of this experiment is to determine if the presentation of a moving auditory signal can affect the perceptual organization of a stroboscopic visual display. A second objective of this experiment is to determine to what extent the perceptual organization can be altered by the presence of the moving auditory stimuli.

This experiment can also be viewed as an investigation of the structure and characteristics of the process model depicted in Figures 3.25. The process model depicted within Figure 3.25, replicated to encompass the two directions of motion being utilized in this experiment, represents apparent motion perception in the horizontal and vertical directions. The combination of the two process models is depicted graphically in Figure 5.3. The perceptual organization is modeled as a comparison of the perceptual strength output by the two process models. The horizontal process model incorporates an auditory influence path because the stroboscopic auditory stimulus is presented contemporaneously with the horizontal visual stimulus. This experiment exercises both temporal and spatial characteristics of the process model as well as involving the perceptual organization process required to combine motion detectors.

The structure of the model depicted in Figure 5.3 supports the investigation of the hypothesis that the stroboscopic auditory stimuli will increase the strength of the perception of horizontal motion. This increase in horizontal motion strength biases the comparison of horizontal to vertical motion strength resulting in a shift of threshold. It is hypothesized that by spatially and temporally linking the auditory stroboscopic display to the horizontal orientation of the visual stimuli, the perception of horizontal motion will dominate the perception of vertical motion due to an increase in the strength of the horizontal motion percept and that the increased horizontally-orientated strength would result in increased reports of horizontal organization.

In Figure 5.3, the strength of the inter-modal horizontal apparent motion perception is represented by  $R_H$  and the strength of the vertical apparent motion perception is represented by  $R_V$ . The auditory influence on  $R_H$  is modeled in Figure 5.3 by the sum of  $H_V$  and  $R_A$ . The inter-stimulus interval is constant throughout the experiment and  $I_V$  is equal to  $I_A$ . The vertical extent,  $V_V$ , is constant throughout the experiment. The visual horizontal extent,  $E_V$ , and the auditory horizontal extent,  $E_A$ , are manipulated in the experiment.

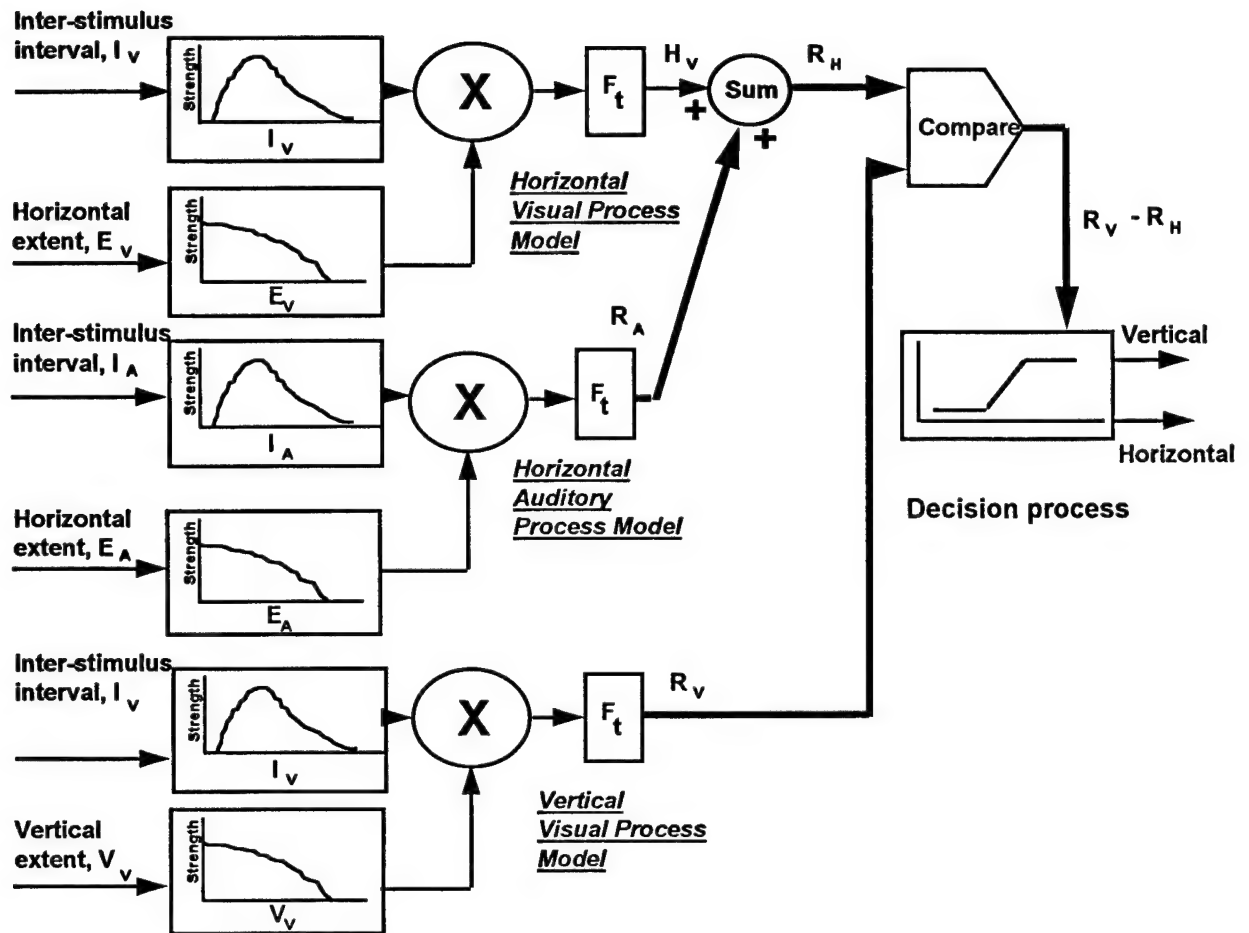


Figure 5.3: The process model depicted in Figure 3.25 is used to form a larger model. This larger model uses two process models, an *inter*-modal model representing horizontal apparent motion perception and an *intra*-modal model representing vertical apparent motion perception. The strengths of the horizontal and vertical apparent motion perceptions are compared to one another. The output of this comparison drives a decision processor.

### 5.1.2 Independent variables

There was a single independent variable, the auditory presentation mode,  $S_{mode}$ , which could take on one of two values, either stationary (S) or switching (M). The auditory signal was presented contemporaneously with the visual stimulus. The auditory signal became audible with the onset of the visual display but did not become inaudible during the ISI. Previous work by Strybel *et al* in auditory apparent motion had shown that auditory presentations that become inaudible during the ISI tended to destroy the illusion of motion created by the rapid changing of auditory source position [120]. This effect could be seen in Figure 3.19, as the sharp peak of this curve was at an ISOI of 50ms with rapid onset under ISOIs less the 50ms and rapid decay under ISOIs greater than 50ms. The duration of the auditory source in Strybel's experiment was 50ms. The visual pattern was not view-able during the ISI.

In the stationary mode ( $S_{mode} = S$ ), the auditory signal was spatially coincident with the lower left visual dot. In the switching mode ( $S_{mode} = M$ ), the auditory signal alternated between the left and right visual dot and was linked with visual presentation by switching positions coinciding temporally with the single visual dot in frame one and then the lower right visual dot in frame two. The auditory stimulus remained in either the left or right position through the duration and ISI of the visual stimulus. A visual representation of the auditory presentation locations is shown in Figure 5.1.

### 5.1.3 Dependent variable

A single dependent variable was defined for this experiment. The dependent variable was the threshold horizontal distance,  $E_H$ , at which the visual perception of motion direction was reported by the subject as having changed.

### 5.1.4 Other conditions

The apparatus used during this experiment was the Speaker/LED apparatus. The vertical separation between the LEDs forming the upper and lower dot positions used in the visual stimulus affects the absolute value of the reported horizontal distance threshold  $E_H$ . The value of vertical separation,  $E_V$ , used in this experiment was  $5^\circ$ .  $E_V$  was held constant by rigidly placing the subject in an established position relative to the visual display and using a chin-rest to maintain the subject in that position. By holding  $E_V$  constant, the ratio of the horizontal extent to vertical extent did not need to be computed for each subject and trial, only  $E_H$  needed to be collected.

### 5.1.5 Subjects

Eight subjects were used in this experiment. Six were male, 2 were female, and all were between the ages of 18 and 35. Each subject reported no known hearing impairments. Each of the subjects had normal or corrected-to-normal visual acuity with one of the male subjects reporting slight color-blindness. Each subject was a volunteer paid for their time as a subject.

### 5.1.6 Task

The method of limits was utilized. Each subject was run in one session that lasted approximately one hour. Each session included a verbal introduction to the overall goals of the research facility, the equipment making up the facility, reading and signing a consent form, training using the experimental apparatus, and finally datum collection. The subjects were naive in terms of the hypothesis of the experiment. Instructions to the subject were standardized and read prior to data collection. These instructions are contained in Appendix B.

The collection of data consisted of 16 trials for each subject. The trials were self-paced and each lasted approximately 90 seconds. Each trial consisted of a series of ascending and descending presentations in which  $E_H$  was modified after each presentation. The initial configuration of the visual stimulus for the ascending trials is shown on the left of Figure 5.2 and the initial configuration of the visual stimulus for the descending trials is shown on the right of Figure 5.2. The subjects were asked to report the direction of the most robust motion, either vertical or horizontal. The end of a series was determined by the subject when they reported a change in the direction of movement from the previous presentation.  $E_H$  was collected at the end of each ascending and descending series. The 16 trials were arranged as eight repetitions of the 2 treatment condition combinations. The order of the treatment combinations was block randomized and different for each subject.

Training using the speaker/LED apparatus was initiated with the subjects being instructed to listen to an audio stimulus as it was presented from the leftmost speaker and then from the rightmost speaker. The spectra of the auditory stimulus had a peak at approximately 400Hz, decreased at approximately 10dB/octave toward lower frequencies, and decreased at approximately 3dB/octave toward the higher frequencies. In addition, at approximately 8000Hz, the spectra began to decrease at approximately 10dB/octave. The subjects were then instructed to listen to the stimulus moving from speaker to speaker horizontally for approximately 30 seconds. Two training trials were then administered with feedback given to the subject. The training concentrated on familiarizing the subject with how to respond using the button response box after each presentation of the stimulus and demonstration of horizontal and vertical visual apparent motion perception. The first datum collection set was then administered.

The stimulus was presented as an alternating series of ascending and descending

trials, beginning with ascending, with  $E_H$  beginning as a small angular extent, ranging from 40% to 70% of  $E_V$ , and slowly increasing for the ascending trials until the subject reported a change in the presentation from horizontal to vertical motion. During the descending trials,  $E_H$  began as a large angular extent, ranging from 140% to 170% of  $E_V$ , and slowly decreased until the subject reported a change from vertical to horizontal motion perception. The auditory and visual stimuli were presented using a duration of 200ms, an ISI of 50ms, and repeated the 1 dot to 2 dot pattern for 7 cycles. After the 7 cycle presentation, the subject was required to choose from either vertical or horizontal motion. The speaker/LED apparatus was used during this experiment and was placed 84 cm (33 inches) from the subjects. The subject's head was stabilized during the experiment using a chin rest that was adjusted for each subject before the experiment began. The visual dots were portrayed by activating one of the green LEDs each with a luminance of  $1.3\text{cd/m}^2$ , and shown against a dark background of  $0.14\text{cd/m}^2$ . Each LED subtended an angle of  $0.31^\circ$  horizontally by  $0.66^\circ$  vertically. The ambient noise level in the experimental area was 54 dB(A) and the auditory level of the noise was 69 dB(A). Sound levels were measured with a B&K Precision Sound Level Meter, type 2235, with a type 4176 microphone insert.  $E_H$  was held constant during the presentation. The spatial extent of the visual stimulus was modified between presentations by the computer using a step size of  $0.33^\circ$ .

After each presentation the subject was required to activate a control indicating which direction, either horizontally or vertically, the stronger perception of motion occurred. No feedback was given to the subjects. Once the subject reported the direction of motion, the visual and auditory stimulus was removed and the sequence was repeated for the next presentation.

### 5.1.7 Results and discussion

The threshold  $E_H$ s obtained from each ascending and descending series were utilized to form a  $RL E_H$  by averaging each pair of  $E_H$  from the ascending and descending series pair. These variables are labeled  $E_{H-asc}$  and  $E_{H-dsc}$  respectively. The mean threshold  $RL E_H$ s are shown on the graphs of Figure 5.4 as are  $E_{H-asc}$  and  $E_{H-dsc}$ .

Analyses of variance were performed on  $E_H$ ,  $E_{H-asc}$ , and  $E_{H-dsc}$  as affected by  $S_{mode}$ . The analyses of variance indicated that neither  $E_H$ ,  $E_{H-asc}$ , or  $E_{H-dsc}$  were significantly affected by  $S_{mode}$ . Summary tables of these analyses of variance are shown in Tables 5.1, 5.2, and 5.3.

The analysis of variance indicates that the reported threshold of the visual motion perception switch from vertical to horizontal or horizontal to vertical movement is not affected by the presence of the moving auditory stimulus. However, the graphs shown in Figure 5.4 depict a trend that is consistent with the hypothesis that the moving auditory stimulus will increase the threshold of the reported visual motion perception switching. Further analyses of the data were performed to illuminate any potential causes of the apparent graphical trend of the data and the statistical

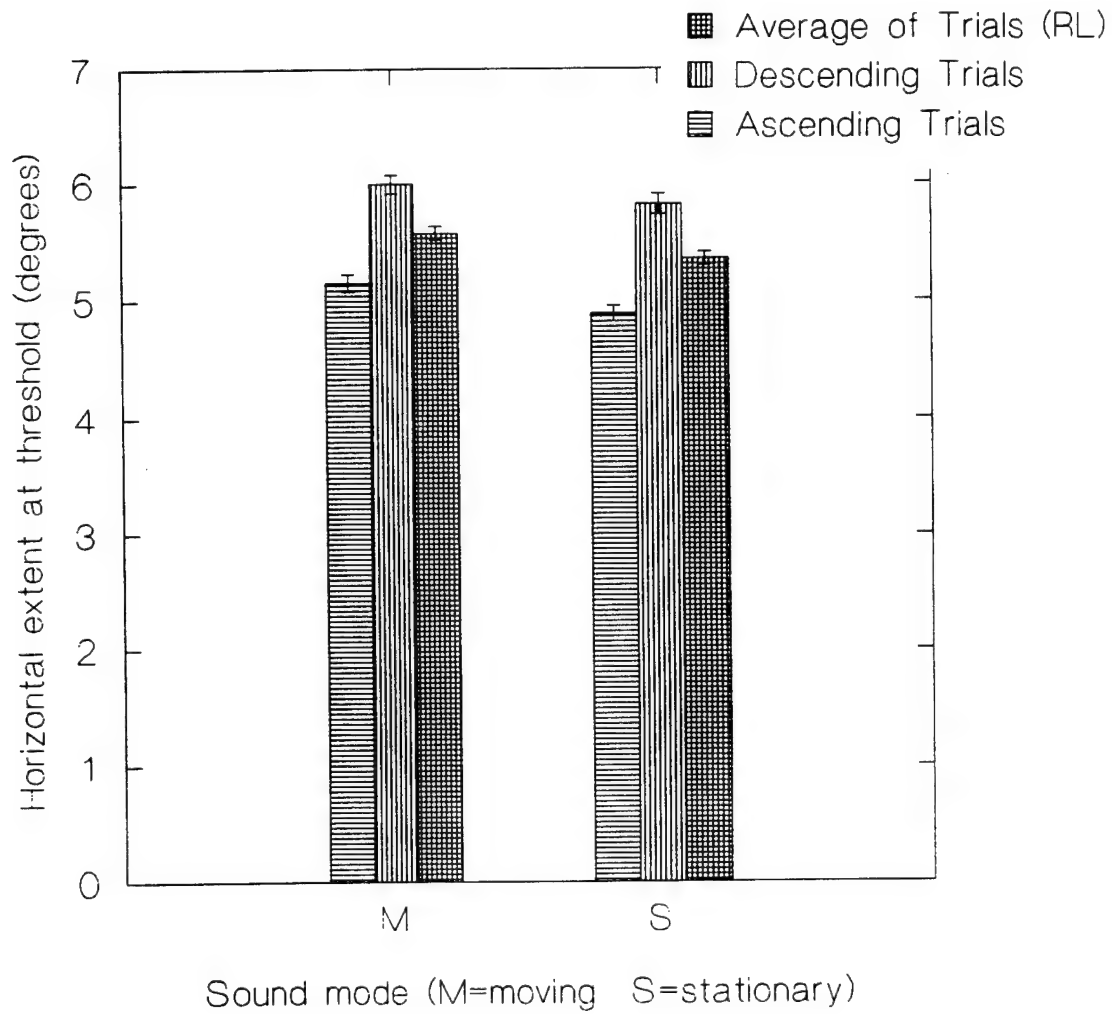


Figure 5.4: The auditory presentation mode did not significantly affect the threshold horizontal extent based on an analysis of variance.

Source	SS	df	MS	F
$S_{mode}$	1.54	1	1.54	3.96
Subject	10.64	7	1.52	
Subject x $S_{mode}$	2.72	7	0.39	

Table 5.1: The averaged threshold  $E_H$ s were not significantly affected by the auditory presentation mode ( $p \geq .05$ ).

Source	SS	df	MS	F
$S_{mode}$	2.20	1	2.20	4.47
Subject	8.31	7	1.19	
Subject x $S_{mode}$	3.44	7	0.49	

Table 5.2: The ascending trial  $E_H$ s were not significantly affected by the auditory presentation mode.

Source	SS	df	MS	F
$S_{mode}$	0.99	1	0.99	2.15
Subject	25.09	7	3.59	
Subject x $S_{mode}$	3.24	7	0.46	

Table 5.3: The descending trial  $E_H$ s were not significantly affected by the auditory presentation mode.

analyses.

The variables  $E_H$ ,  $E_{H-asc}$ , or  $E_{H-dsc}$  are shown in Figures 5.5, 5.6, and 5.7 respectively, plotted as a function of another variable, t-block. T-block was a temporal indicator that combined data from 4 sequential trials into a single set. In other words,  $E_H$ , when t-block was 1, was the mean of  $E_H$  over the first 4 trials. The second set of 4 trials was averaged for t-block of 2, and so on. Thus, t-block ranged from 1 to 4.

Figures 5.5, 5.6, and 5.7, begin to show a possible explanation for the lack of statistical significance in the previous analyses of variance. These graphs begin to show that the influence of the moving auditory stimulus may not have been constant over the length of the experimental period. Specifically, the graphs indicate that when t-block is 2, the maximum influence of the moving auditory stimulus over the visual perception may have occurred.

A set of paired t-tests was performed to verify this second hypothesis. The analysis of this second hypothesis is a post-hoc comparison. Special attention must be given to post-hoc comparisons in that they will increase the probability of rejecting the null hypothesis when it is actually true. However, post-hoc comparisons are extremely useful in exploratory data analyses which provide guidance for future experiments. To this end, a t-test was performed for each pair of datum sets at each t-block value, checking for the possibility that the means of  $E_H$ ,  $E_{H-asc}$ , or  $E_{H-dsc}$  were different under the two conditions of  $S_{mode}$ .

The results of the paired t-tests indicate that the difference between the means was not significantly different. However, this may have occurred due to several reasons, one being that for each subject, only two datum points were available for



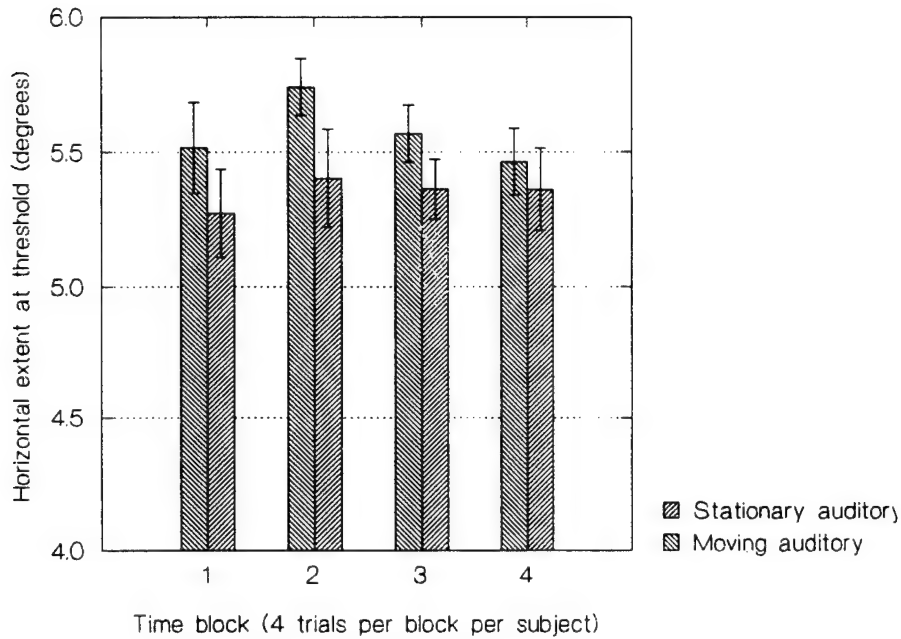


Figure 5.5: Results of a paired t-test indicated that order did not affect the averaged threshold extents when the data were blocked into four temporal groups.

each value of  $S_{mode}$  and each value of t-block.

Overall, the statistical analyses performed within this experiment have provided conflicting results with the graphical trends in the data. There are potentially several causes for these conflicts.

One potential contributor to the conflicting trend and statistics may have been that not enough subjects were utilized in the study. If the numbers of subjects were increased, the error variance for the within-subject testing might decrease and thus increase the separation between the datum sets statistically. The thresholds obtained from the ascending and descending trials did vary among subjects. To illustrate this, graphs depicting  $E_H$ ,  $E_{H-asc}$ , and  $E_{H-dsc}$  as functions of subject, for the moving and stationary auditory presentation conditions, are shown as Figures 5.8, and 5.9.

By combining the data from Figures 5.8 and 5.9 for  $E_H$ , on one graph, shown as Figure 5.10, it can be seen that only six of the eight subjects show the tendency of an increase in threshold due to the presence of moving auditory stimuli as compared

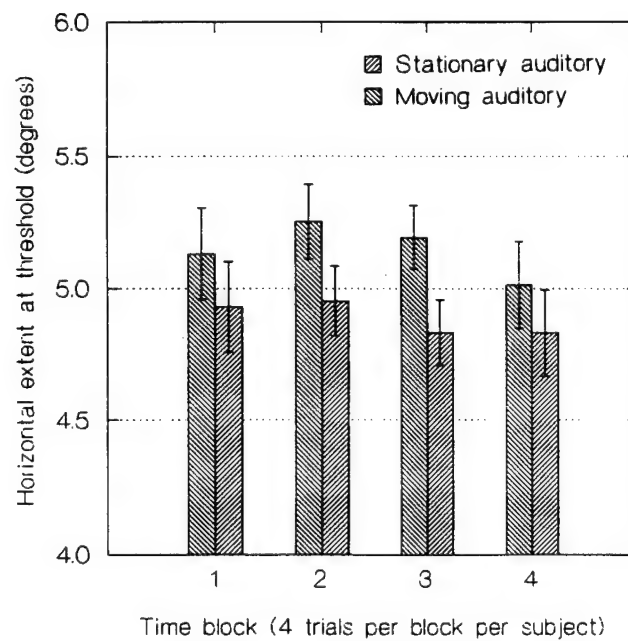


Figure 5.6: Results of a paired t-test indicated that time did not affect the threshold extents of the ascending trials when the data were blocked into four temporal groups.

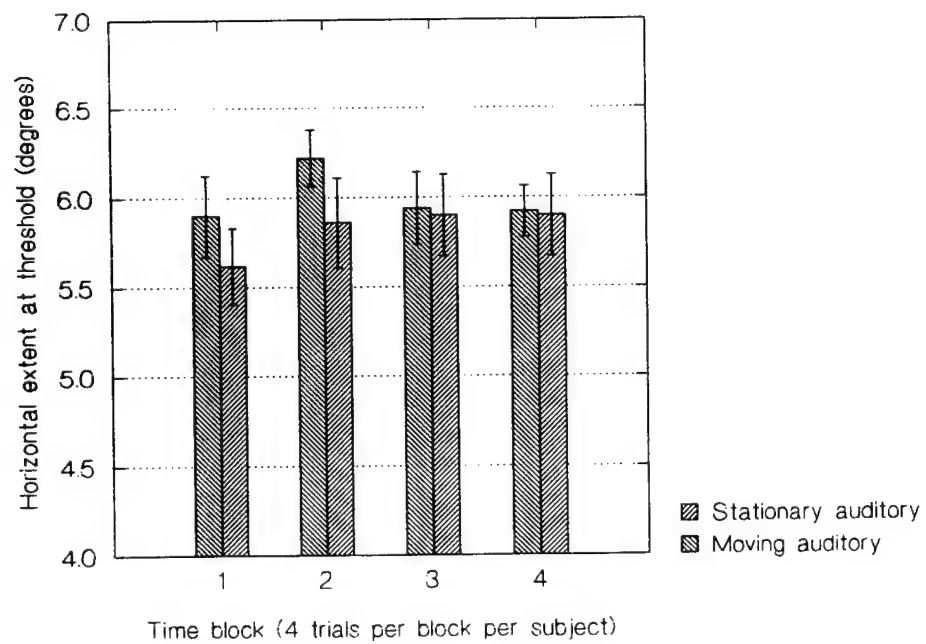


Figure 5.7: Results of a paired t-test indicated that time did not affect the threshold extents of the descending trials when the data were blocked into four temporal groups.

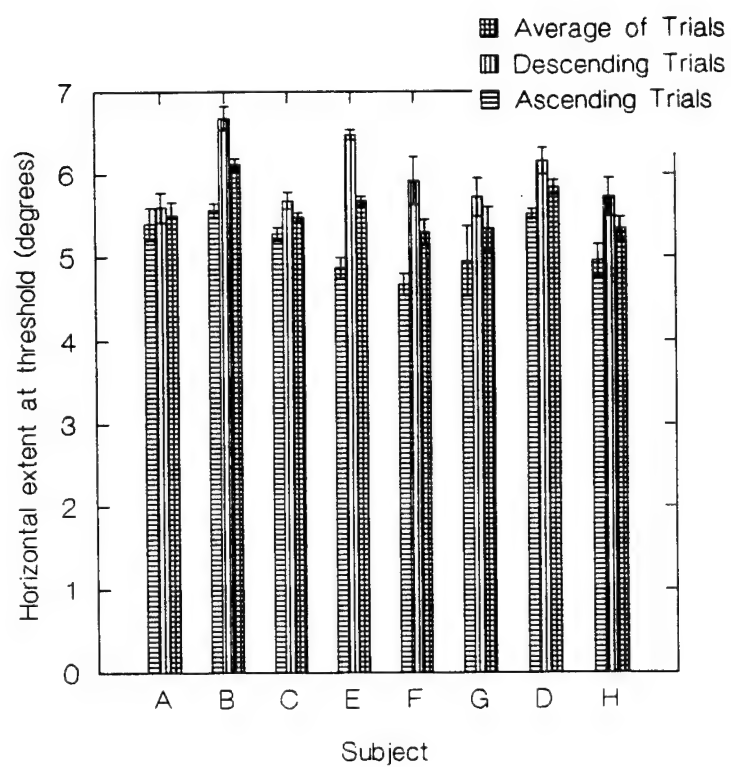


Figure 5.8: The thresholds obtained from the moving auditory presentation trials varied between subjects.

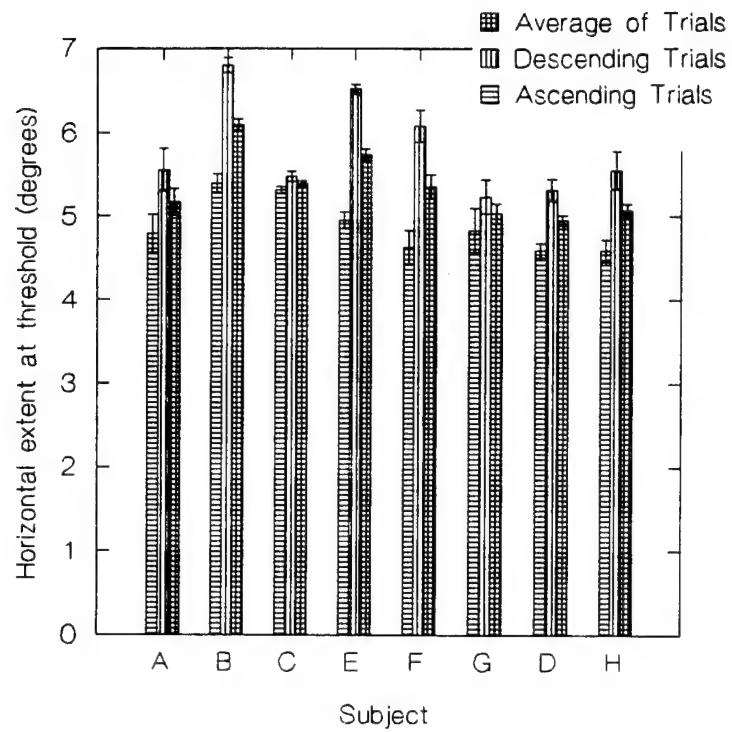


Figure 5.9: The thresholds obtained from the stationary auditory presentation trials varied between subjects

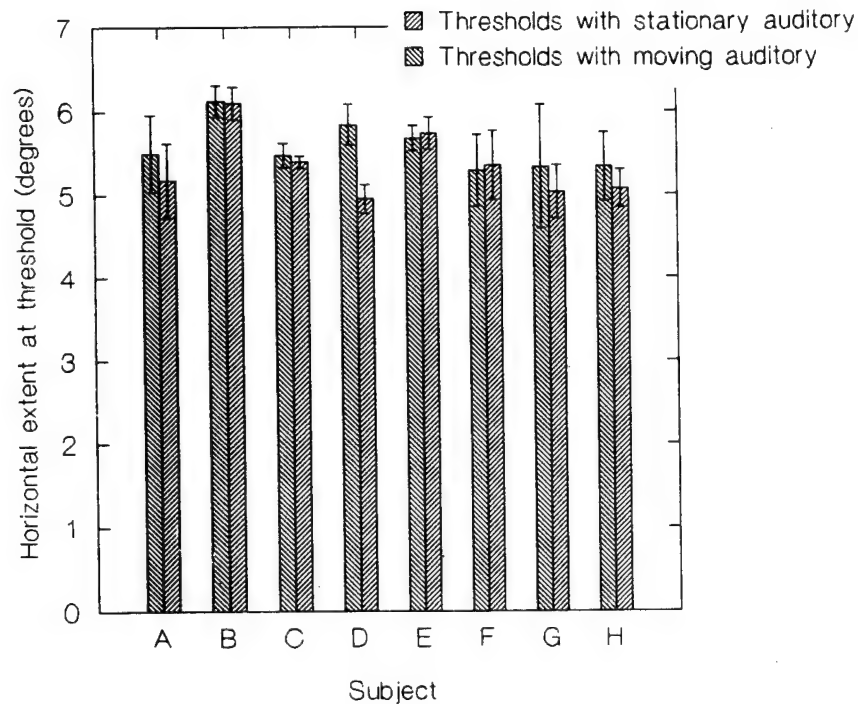


Figure 5.10: Six of the eight subjects showed the tendency of an increase in threshold due to the presence of moving auditory stimuli as compared to the threshold obtained with stationary auditory stimuli.

to the threshold obtained with stationary auditory stimuli. In addition, the effect appears to be weak in the six subjects that do exhibit the increase in threshold means. Because of these facts, it is unlikely that increasing the number of subjects would affect the statistical significance of the analyses of variance.

Another potential contributor could have been that a varying threshold was utilized by each of the subjects to trigger the decision to report a change in motion direction. This would have manifested itself first within each of the t-blocks as a larger than expected variance on each of the t-block means. This increased variance would carry through the calculation of overall means and, while not necessarily affecting the mean itself, might increase the variance of the overall mean. This appears somewhat unlikely in that the variance for each subject in the ascending and descending trials appears to be reasonably consistent as seen in Figures 5.8 and 5.9.

Another potential contributing factor may have been the experimental apparatus itself. The apparatus used in this experiment did not allow the moving sound source

to be accurately linked spatially with the light bar LEDs because the speakers were separated by a fixed distance and rigidly mounted to the light bar. The lack of accurate spatial linkage may have enhanced, reduced, or modified the influence of the auditory signal on the visual perception. This potential is reinforced by the fact that the minimum spatial auditory resolution reported in the literature is between  $1^\circ$  to  $2^\circ$ , which is less than the spatial linkage disparity between many of the LEDs and the speakers in the experimental apparatus.

To investigate if the lack of accurate spatial linkage between the visual and auditory sources may have affected the results of this experiment, a virtual auditory display generator could be utilized which would be more capable of maintaining spatial and temporal linkage between the visual and auditory stimuli than the speaker/LED apparatus. In addition, the use of a virtual auditory display would enable more complex movement patterns to be generated when compared to the use of a conventional speaker mounted on a moving platform.

While the benefits to be derived from the use of a virtual auditory display generator are clear, the virtual nature of the auditory generator brings with it some attributes which may adversely affect the ability to generalize any empirical data obtained using the device onto other conditions. This would occur if the results obtained were significantly affected by some characteristic peculiar to the device itself.

The capability of the virtual auditory display generator, or auditory localizer as it is sometimes called, to provide moving auditory stimuli equivalent to conventional moving auditory stimuli, as perceived by the human, over the range of use considered for the device, must be assessed to determine if the experimental results obtained with the device are general in nature. Because stroboscopic auditory displays are utilized within this research, both static auditory localization and dynamic auditory image perceptual equivalence must be assessed. To evaluate the static localization capability and dynamic characteristics of the localizer, the pertinent literature was reviewed and two experiments with associated analyses were performed. This evaluation is detailed in the following chapter.

## Chapter 6

# Localizer dynamic evaluation

The static localization capability and dynamic characteristics of the specific virtual auditory display generator, or auditory localizer, used within this research program was evaluated through a review of pertinent literature and a set of experimental studies. The virtual localizer was integrated within the Localizer/CRT apparatus as discussed in Chapter 4. The results of the static and dynamic evaluation of this particular auditory localizer are presented in this chapter.

This evaluation was necessary to assess the generalized nature of empirical results obtained using this localizer as part of the auditory stimuli. This auditory localizer must provide moving and static auditory stimuli perceptually equivalent to conventional moving or static auditory stimuli if the empirical results are to be generalized to other auditory display conditions. There were no known procedures to accomplish this evaluation and because of this, a novel set of tests were applied.

### 6.1 Localizability of static stimuli

The quality of the auditory stimuli had to be assessed before it could be assumed that the results obtained in experiments using the Localizer/CRT apparatus could be generalized to other environments. Because the visual display was real (not virtual) in nature, the localizability of visual stimuli was not in question. However, of more concern was the auditory localizer, which was a novel virtual device that provided a silicon-based simulation, using digital signal processing algorithms, of an auditory source emitting in an anechoic chamber.

One particular aspect of this localizer which made this assessment necessary was the fact that individualized pinna models were not used in this localizer. By not using individual pinna models, optimal performance of the localizer might not be achieved. However, if equivalent perceptual performance could be maintained without the use of individualized pinna models over the limited movement characteristics employed within the research, empirical results obtained using the device could still be generalized across other environments.



Audio source	Localizer Mean error	Free field Mean error
pink noise	4.8°	5.5°
male speech	4.8°	6.0°
female speech	4.7°	6.1°
250Hz tone	5.2°	5.7°
500Hz tone	4.8°	6.0°
1kHz tone	4.6°	6.0°
2kHz tone	4.6°	6.9°
4kHz tone	5.9°	6.5°
8kHz tone	4.5°	6.5°

Table 6.1: A summary of Erikson and McKinley's testing of static localization errors indicated that the localizer provided an equivalent auditory environment referenced to a single sound source in a free field [80].

Researchers have compared the static performance of this particular localizer design to real sound sources in anechoic chambers in terms of static localization accuracy in past studies [80]. The results of these comparisons are found documented in the literature and are reviewed below.

The results of previous static testing indicated that this device elicits localization accuracy equivalent to real static sound sources in anechoic chambers in azimuth [80], [28]. In particular, Erikson and McKinley evaluated static localization performance relative to localization in a free-field anechoic chamber using 10 subjects under several audio sources, including octave-band noise, broad-band noise, male speech, and female speech [28]. In this study, the subjects were asked to point their heads in the direction of the sound source. The head pointing angle was measured and the difference between the angle of the auditory stimulus and the subject's head angle was defined as the directional error. Each subject used in this study was found acceptable using an audio-metric test from 500Hz to 8kHz. A summary of Erikson and McKinley's results, taken from [80], is shown in Table 6.1. Individual subject means of directional error ranged from 2.3° to 13.0° [28].

Overall, the mean directional errors found in this study in the free field anechoic chamber were 6.03° with a standard deviation of 1.3° compared with a mean directional error using the localizer of 4.84° with a standard deviation of 1.6° [28]. Erikson and McKinley stated that testing of static localization errors indicated that the localizer provided an equivalent auditory environment referenced to a single sound source in a free field [28] [80].

Other researchers have also evaluated the static performance of localizers incorporating different designs but also not using individualized pinna models. Wightman performed a comparison of auditory localization in a free-field anechoic

chamber and using an auditory localizer system [140]. In this study, Wightman presented eight subjects with auditory stimuli that varied in azimuth and elevation and asked the subjects to orally report the position of the sound source in degrees. Wightman used 10 hours of practice per subject to reduce the variance caused by the verbal reporting method. Wightman stated that localization in azimuth was equivalent in both environments with a mean error in the free-field being  $19.4^\circ$  and the mean error of the localizer being  $17.8^\circ$  [140]. Many differences appear to exist in the experimental apparatus and methodologies used by researchers employing auditory localizers. Differences such as which head related transfer functions were used, the extent of azimuth and elevation angular coverage manipulated, and the digital processing algorithms used, may have altered the relative accuracy reported in these experiments.

Other empirical efforts have provided similar claims for other auditory localizers in azimuth [141]. Wightman also found that the sensitivity to filter characteristics of individual auditory systems, such as pinnae, canals, and head-forms, have a negligible effect on azimuth but a profound effect on elevation localization [141].

The Wightman, Erickson, and McKinley studies all support the conclusion that in azimuth, this particular localizer is perceptually equivalent to a single static sound source emanating within an anechoic chamber.

Verification that the specific localizer embedded in the Localizer/CRT apparatus was operating in a manner equivalent to the particular localizer utilized by McKinley [80], and Erikson and McKinley [28], was performed. The verification consisted of an analysis of the power spectral density of the output of the localizer when driven by a white noise source.

The power spectral density was obtained using a Lecroy sampling oscilloscope and a B&K audio noise generator, Type 1405. Samples of the localizer output power spectral densities at several input angles are shown in Figure 6.1. The PSDs obtained from this testing were similar to those documented by McKinley [80]. This was not unexpected in that the localizer integrated into the Localizer/CRT apparatus was designed and fabricated by McKinley, and the software load was of the same configuration as used in previous studies by McKinley. This comparison verified that the auditory localizer within the Localizer/CRT apparatus was operating properly.

While previous testing of this localizer had been limited to static stimuli, it remained necessary to evaluate the localizer's dynamic characteristics. The evaluation of the dynamic characteristics of this, or any other, auditory localizer had not been documented in the literature.

The evaluation of the dynamic characteristics of this particular localizer was pursued along a psychophysical axis, which assessed the ability of the device to elicit equivalent perceptual results found with non-virtual devices. The auditory display was evaluated in terms of its ability to produce auditory stimuli that supported acceptable discrimination of velocity and dynamic resolution. These empirical evaluations are documented in the following sections.

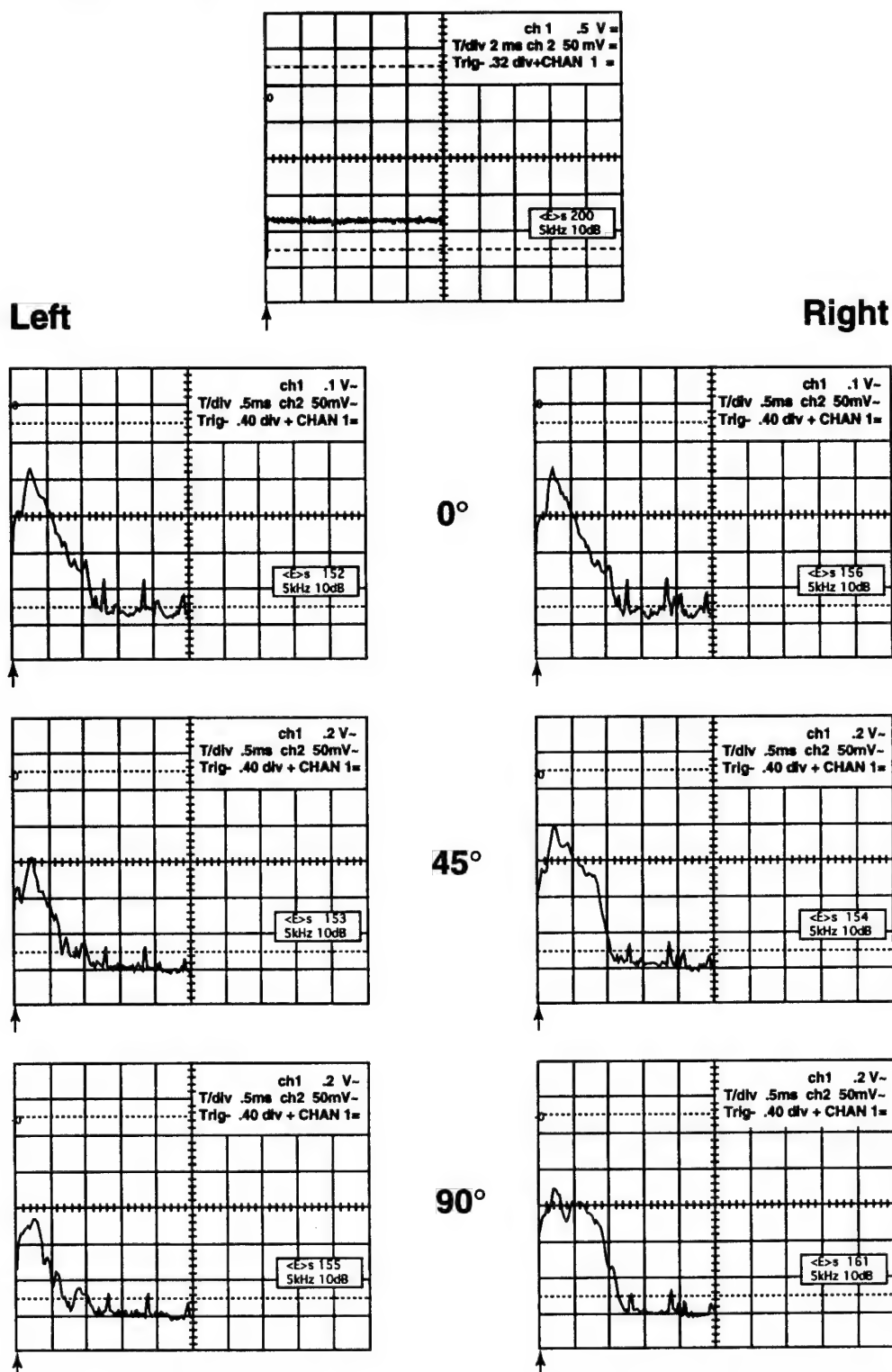


Figure 6.1: Power spectral density measurements of the localizer integrated into the Localizer/CRT apparatus, at several different angles, were similar to those documented by McKinley [80]. Boldface numbers in the center of the figure are azimuth angles.

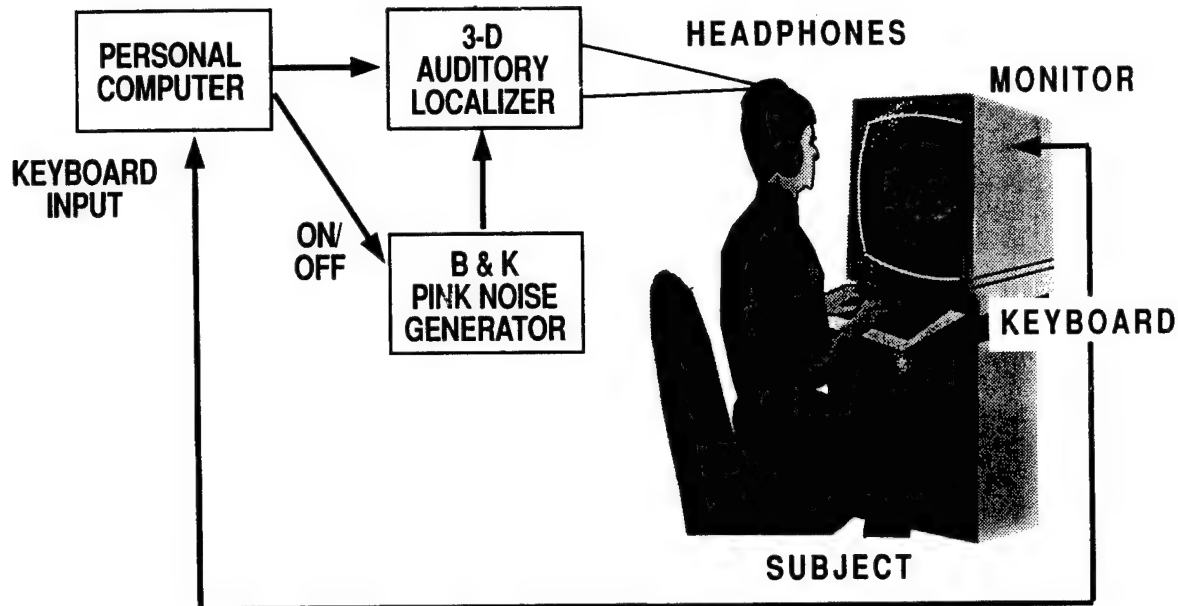


Figure 6.2: In experiment two, the subject was seated at a desk and listened to moving auditory presentations of constant angular velocity with respect to the subject.

## 6.2 Experiment two: Auditory velocity discrimination

Auditory velocity discriminability in this experiment is defined as the ability of a subject to reliably report that a velocity difference exists between two sequentially presented moving sounds. The ability of humans to discriminate velocity is documented in existing literature but the level at which velocity discrimination can occur differs from study to study. This experiment empirically quantified a velocity discrimination level using this virtual auditory display.

A subset of the Localizer/CRT station equipment was utilized to conduct this experiment. Specifically, the graphics generator, graphics monitor, HP signal generator, audio mixer, audio filter, noise gate, and head tracker were not used. A block diagram of the experimental apparatus used in this experiment is shown in Figure 6.2.

### 6.2.1 Aims

The objective of this experiment was to determine the quality of velocity discrimination of human observers using a particular auditory virtual environment generator.

### 6.2.2 Independent variables

Six levels of velocity difference( $\Delta_v$ ) ( $2^\circ/\text{sec}$ ,  $5^\circ/\text{sec}$ ,  $8^\circ/\text{sec}$ ,  $11^\circ/\text{sec}$ ,  $14^\circ/\text{sec}$ ,  $17^\circ/\text{sec}$ ), and four horizontal extents, or angular amplitudes, ( $E$ ) were chosen to move the sound source through ( $\pm 10^\circ$ ,  $\pm 40^\circ$ ,  $\pm 90^\circ$ ,  $\pm 130^\circ$ ).

### 6.2.3 Other conditions

Three *base* velocities were chosen over which the  $\Delta_v$  was applied. The base velocities were  $20^\circ/\text{sec}$ ,  $60^\circ/\text{sec}$ ,  $100^\circ/\text{sec}$ . The faster presentation was either the first or second presentation, which doubled the number of trials for each subject. The combinations of  $E$ ,  $\Delta_v$ , and *base* velocity, yielded 144 treatments conditions. The order of the 144 treatments was randomized for each subject. While the *base* velocity was not of primary interest, it was randomly applied across subjects. The audio signal was generated by a B&K audio noise generator, Type 1405. The spectra of the audio signal had a peak at approximately 80Hz, decreased at approximately 6dB/octave toward lower frequencies, and decreased at approximately 3dB/octave toward the higher frequencies. In addition, at approximately 21kHz, the spectra began to decrease at approximately 25dB/octave. The sound pressure level was adjusted at the beginning of each session by each subject to a comfortable level.

### 6.2.4 Subjects

Six male volunteer subjects were utilized with the author being one of the six. None of the subjects had any known hearing impairments.

### 6.2.5 Task

The experiment was conducted using a two-alternative forced-choice paradigm. Each subject was exposed to all treatment combinations. During each trial, the subject was prompted that the first auditory presentation was to begin. The presentation began to the right of the subject traveling in an arc of constant distance through the median plan ending on the left of the subject, at which point the stimulus was turned off. The second presentation followed the first after approximately 2 seconds. The extent of the second presentation was the same as the first but the velocity was altered. The subject was then asked to press a key indicating which presentation was faster. The subject could take as long as they wished to respond. The next trial began approximately 2 seconds after the subject responded on the keyboard. The sound sources were stabilized to the subject and not to the external environment.

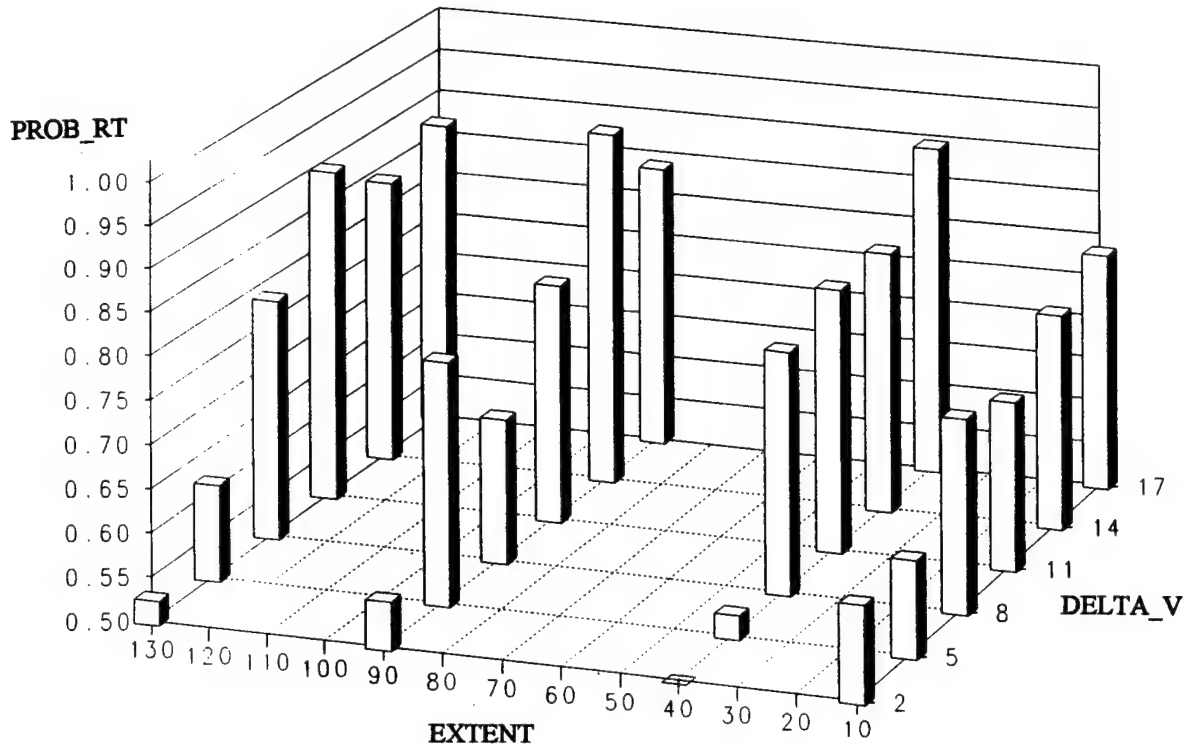


Figure 6.3: The velocity difference between each presentation in a pair, and the horizontal extent of each presentation pair, were independent variables. The mean percentage of correct responses at each level of velocity difference and horizontal extent is plotted.

### 6.2.6 Dependent variable

The percentage of responses which correctly identified the faster velocity was calculated for each combination of treatment levels. This formed the dependent variable represented by  $P_c$ .

### 6.2.7 Results and discussion

A graphical depiction of the probability of being correct, calculated as  $\frac{P_c}{100}$  and represented by PROB\_RT, obtained across all levels of extent, labeled as EXTENT, and velocity differences ( $\Delta_v$ ) labeled as DELTA\_V, is shown in Figure 6.3.

Shown in Figure 6.4 are the data from Figure 6.3 smoothed using the distance weighted least squares (DWLS) algorithm described in the SYSTAT statistical analysis software package [124]. The smoothing obtained a surface reflecting the

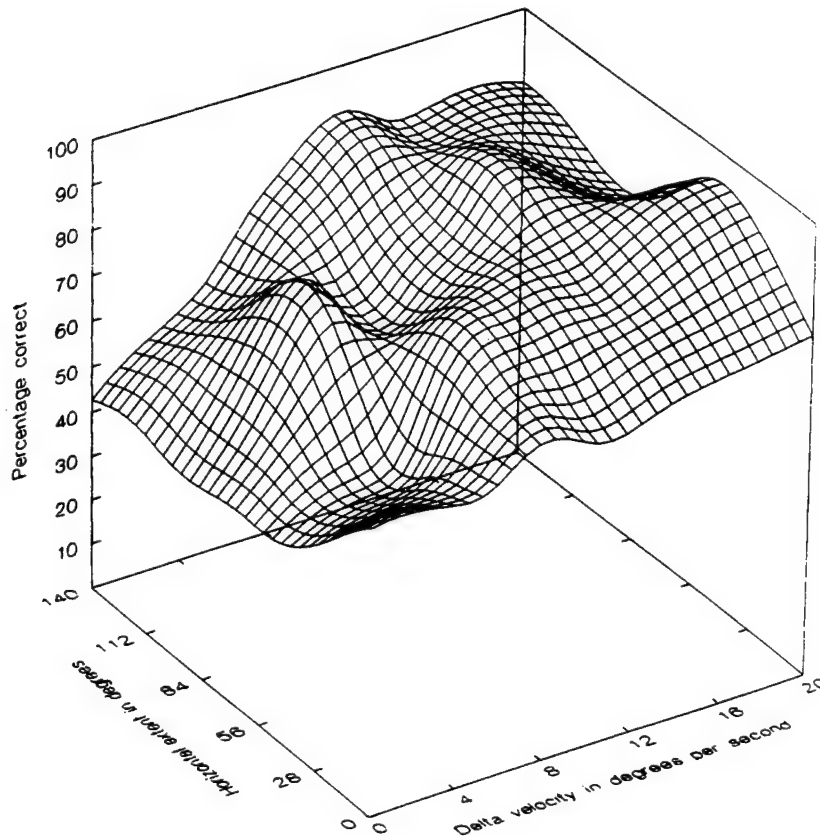


Figure 6.4: The empirical data depicted in Figure 6.3 were smoothed to form a mesh plot.

percentage correct as a function of extent and delta velocity.

An analysis of variance was performed using the probability of correct response, computed as  $(P_c/100)$ , as the dependent variable. Statistically, the effect of extent ( $E$ ) on the probability of correct response was not significant while the difference velocity ( $\Delta_v$ ) ( $F(5, 25) = 11.02, p < .05$ ) and base velocity ( $F(2, 10) = 18.65, p < .05$ ), were significant. No third or second order interactions were statistically significant. Data from the author did not appear to differ from data obtained from the other subjects and the overall variance attributable to subjects was relatively small. Figures 6.5, 6.6, and 6.7 graph  $P_c$  as function of the independent variables and the base velocity. A summary of the analysis of variance results is shown in Table 6.2.

The graph relating the percentage correct with the difference velocity, Figure 6.7, shows a 75% threshold value of approximately  $8^\circ/\text{sec}$  with  $2^\circ/\text{sec}$  clearly being perceived at only a chance level. This could only be viewed as a crude estimate of the threshold value. To refine this estimate further, analyses of the percentage correct data were performed as described in Woodworth and Schlosberg's *Experimental Psychology* [63] as threshold determinations using the method of constant stimuli. In

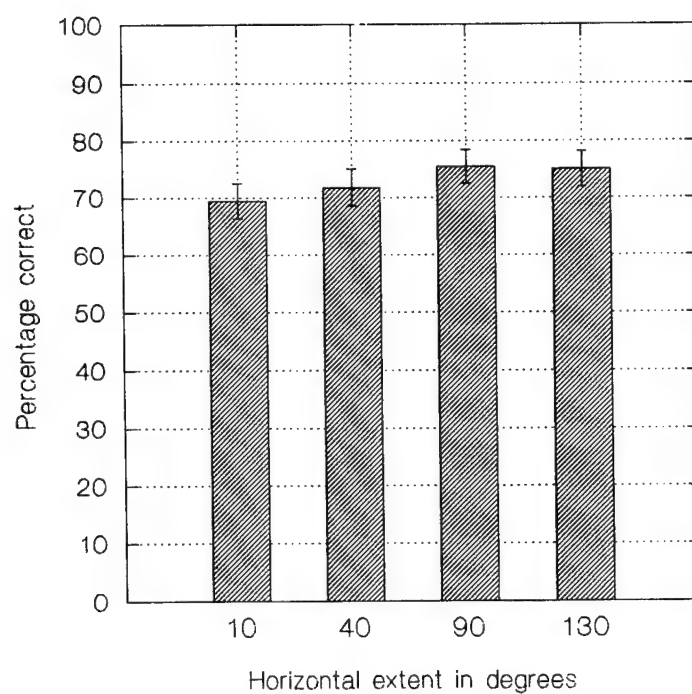


Figure 6.5: The ability of the subject to discriminate auditory velocity was significantly affected by the difference velocity between two presentations but not by the horizontal extent of the auditory presentations.



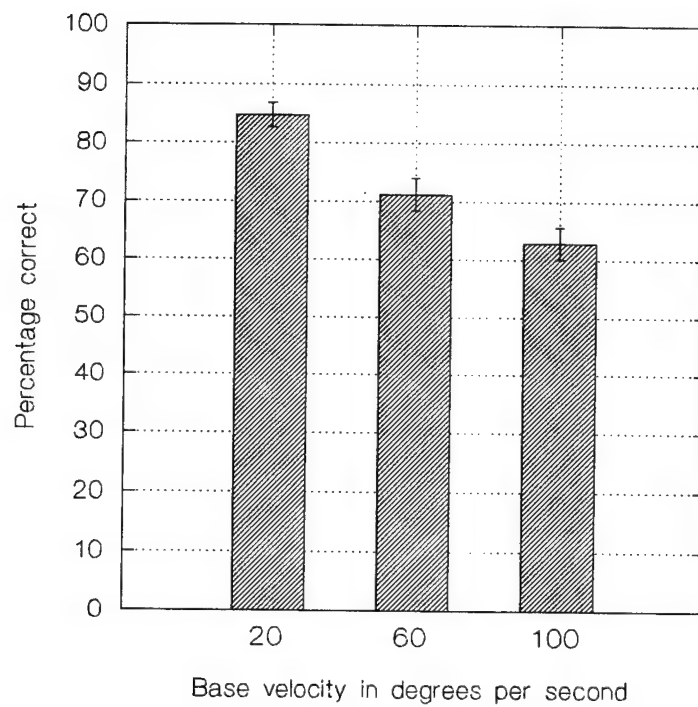


Figure 6.6: Three *base* velocities were chosen over which the  $\Delta_v$  was applied. Base velocity significantly affected velocity discrimination of the auditory presentations.

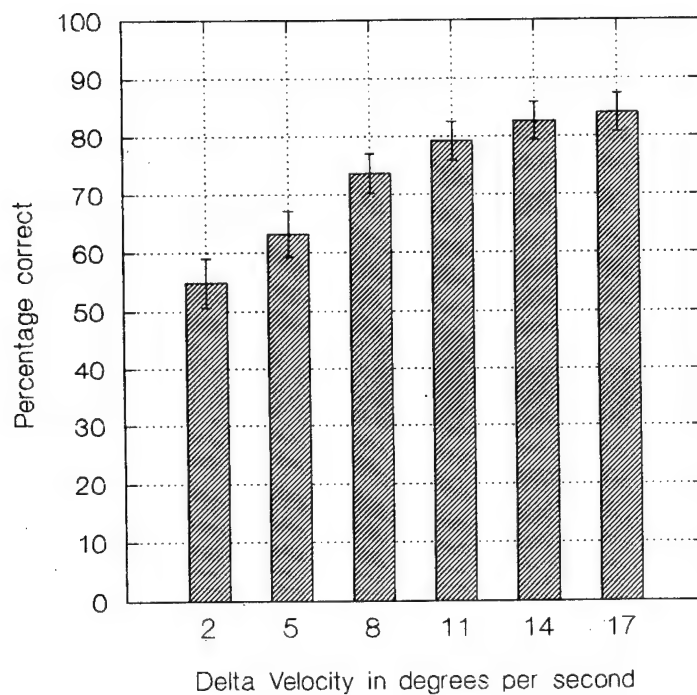


Figure 6.7: The ability of the subject to discriminate auditory velocity was significantly affected by the difference velocity between two presentations but not by the horizontal extent of the auditory presentations.

Source	SS	df	MS	F
$E$	0.52	3	0.17	1.25
Subject	2.68	5	.54	
Subject x $E$	2.09	15	0.14	
$\Delta_v$	9.76	5	1.95	11.02*
Subject x $\Delta_v$	4.43	25	0.18	
Base Velocity	7.02	2	3.51	18.65*
Subject x Base Velocity	1.88	10	0.19	
$E$ x $\Delta_v$	2.67	15	0.18	1.02
Subject x $E$ x $\Delta_v$	13.13	75	0.18	
$E$ x Base Velocity	1.61	6	0.27	1.67
Subject x $E$ x Base Velocity	4.82	30	0.16	
Base Velocity x $\Delta_v$	2.42	10	0.24	1.89
Subject x Base Velocity x $\Delta_v$	6.42	50	0.13	
$E$ x $\Delta_v$ x Base Velocity	3.95	30	0.13	0.75
Subject x $E$ x $\Delta_v$ x Base Velocity	26.21	150	0.17	

\* Significant  $p < .05$

Table 6.2: An analysis of variance was performed to determine if the angular extent, velocity difference, or base velocity significantly affected  $P_c$ . No interactions of the independent variables were indicated but both the velocity difference and the base velocity significantly affected  $P_c$ .

this method, it was assumed that the dependent variable, percentage correct, would follow an ogive curve based on the independent variable, delta velocity, as it progressed from a low value to a high value. As could be seen from the plot of percentage correct versus delta velocity, the data from this experiment did resemble this shape. From these data, z-scores were calculated from the percentage correct, using the mean and variance of the data, and a linear regression was performed on the z-scores. The resultant line was used to find the threshold by finding the z-score at the percentage correct threshold, in this case .75, and then using the straight line to find the value of delta velocity corresponding to that particular z-score. Using this method, the z-score versus delta velocity line had the following form:

$$z = .145\Delta_v - 1.374 \quad (6.1)$$

The datum mean was .754 and the standard deviation was 0.16. That resulted in a calculated z-score at threshold of -.025 and a  $\Delta_v$  of  $9.3^\circ/\text{sec}$ .

Comparison of a calculated Weber fraction, based on the threshold of  $9.3^\circ/\text{sec}$ , to Weber fractions found in the literature, enables analysis of one metric of the localizer quality to be assessed. To quantify a single number representing the percentage velocity required to meet threshold in this experiment, a single value for the base velocity and corresponding threshold velocity must be obtained.

Figures 6.8 and 6.6 show that in this experiment both base velocity and difference velocity affected the threshold values of the probability of a correct discrimination. In Figure 6.8, the probability of being correct, calculated as  $\frac{P_c}{100}$  is labeled as PROB\_RT, obtained across all levels of base velocity, labeled as VEL\_1, and velocity differences ( $\Delta_v$ ) are labeled as DELTA\_V.

The effect of base velocity on the amount of difference velocity required to change the probability of velocity discrimination, seen in Figures 6.8 and 6.6, is consistent with the literature regarding auditory velocity discrimination. To graphically illustrate this effect within this experiment, the data from Figure 6.8 were smoothed using the distance weighted least squares (DWLS) algorithm described in the SYSTAT statistical analysis software package [124] to obtain a surface reflecting the percentage correct as a function of base velocity and delta velocity. This surface is shown as Figure 6.9.

Because three base velocities were utilized equally often, the average base velocity was used for this calculation. The 75% threshold for  $\Delta_v$  was established at  $9.3^\circ/\text{sec}$ . The average base velocity was  $60^\circ/\text{sec}$ . Thus, an estimate for the percentage velocity to meet threshold was  $9.3/60.0$ , or 14.1%.

Four references exist in the literature regarding Weber fractions for velocity discriminations that may form a comparison basis for the Weber fraction of 14.1% found in this experiment. These four references are described and discussed in the following paragraphs.

**Reference 1:** One source of a reference Weber fraction was found in Boff,

Kauffman, and Thomas [12], where the minimum Weber fraction for auditory motion perception, after detection of motion, was stated at 0.05, or 5%. This figure was derived by Boff from a reference by Perrott [34] [92]. This figure was described by Perrott as a difference Limen obtained in an experiment using the method of adjustment that tested the ability of subjects to adjust the velocity of a rotating sound source in an anechoic chamber to match an orally stated reference velocity in miles per hour.

**Reference 2:** In a series of visual angular velocity discrimination experiments, Kaiser found Weber fractions for visual angular rate discrimination from 8% to 20% [60]. The stimuli used in these experiments were simultaneously presented visual objects rotating about their own axes. The observer was asked to indicate which of the two presentations was rotating faster.

**Reference 3:** Altman and Viskov found Weber fractions to be a linear function of the reference velocity using dichotic click-trains [2]. Altman and Viskov found the  $DL$  to be  $10.8^\circ \text{sec}^{-1}$  at  $14^\circ \text{sec}^{-1}$  up to  $19.3^\circ \text{sec}^{-1}$  at  $140^\circ \text{sec}^{-1}$  [2]. This corresponds to Weber fractions ranging from 14% to 77% and can be interpolated to a reference velocity of  $60^\circ \text{sec}^{-1}$  resulting in a velocity discrimination of  $13.9^\circ \text{sec}^{-1}$  or approximately 23%.

**Reference 4:** In a study utilizing a similar methodology to the methodology of this experiment, Grantham found a difference velocity of  $7.5^\circ \text{sec}^{-1}$  using a base velocity of  $40^\circ \text{sec}^{-1}$  resulting in a difference Limen of approximately 19% [39]. Grantham also found that the difference Limen would significantly increase if the signal durations were less than 300ms [39].

The Weber fraction estimate obtained within this experiment of 14.1% corresponds well with results for velocity discriminations found by Kaiser in the visual domain as well as results documented in studies performed by Altman and Viskov in 1977 and Grantham in 1986 for simulated auditory motion. However, it is roughly three times the minimum Weber fraction determined by Perrott [92]. The increased threshold in velocity difference found in this experiment over the minimum threshold reported by Boff, Kauffman, and Thomas [12], as a reference to Perrott [92], may have resulted from differences in the experimental design and datum analysis techniques utilized within this experiment relative to those utilized by Perrott.

The 5%  $DL$  was obtained by Perrott in an experiment utilizing the method of adjustment using three subjects [92]. Perrott evaluated how well a subject could match a rotating sound source velocity to a verbally-presented linear velocity expressed as miles-per-hour. In Perrott's study, no auditory match was given to the subject to establish a correspondence to the verbally- presented mile per hour rate. Specifically, in Perrott's study, each subject's association between the speed of the moving sound source, in degrees per second, and the verbally-stated miles-per-hour, was established as an individual relationship. Using this relationship, an individualized mile-per-hour speed was associated with a  $100^\circ \text{sec}^{-1}$  moving sound

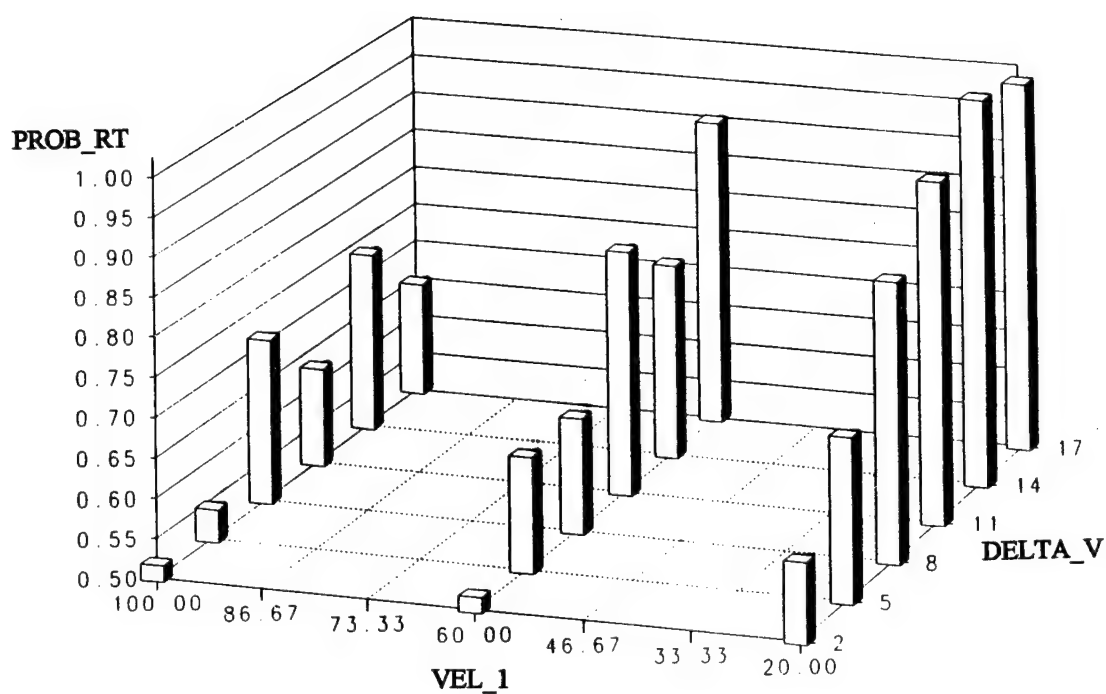


Figure 6.8: Base velocity, shown as VEL\_1, and delta velocity, shown as DELTA\_V, significantly affected velocity discrimination.

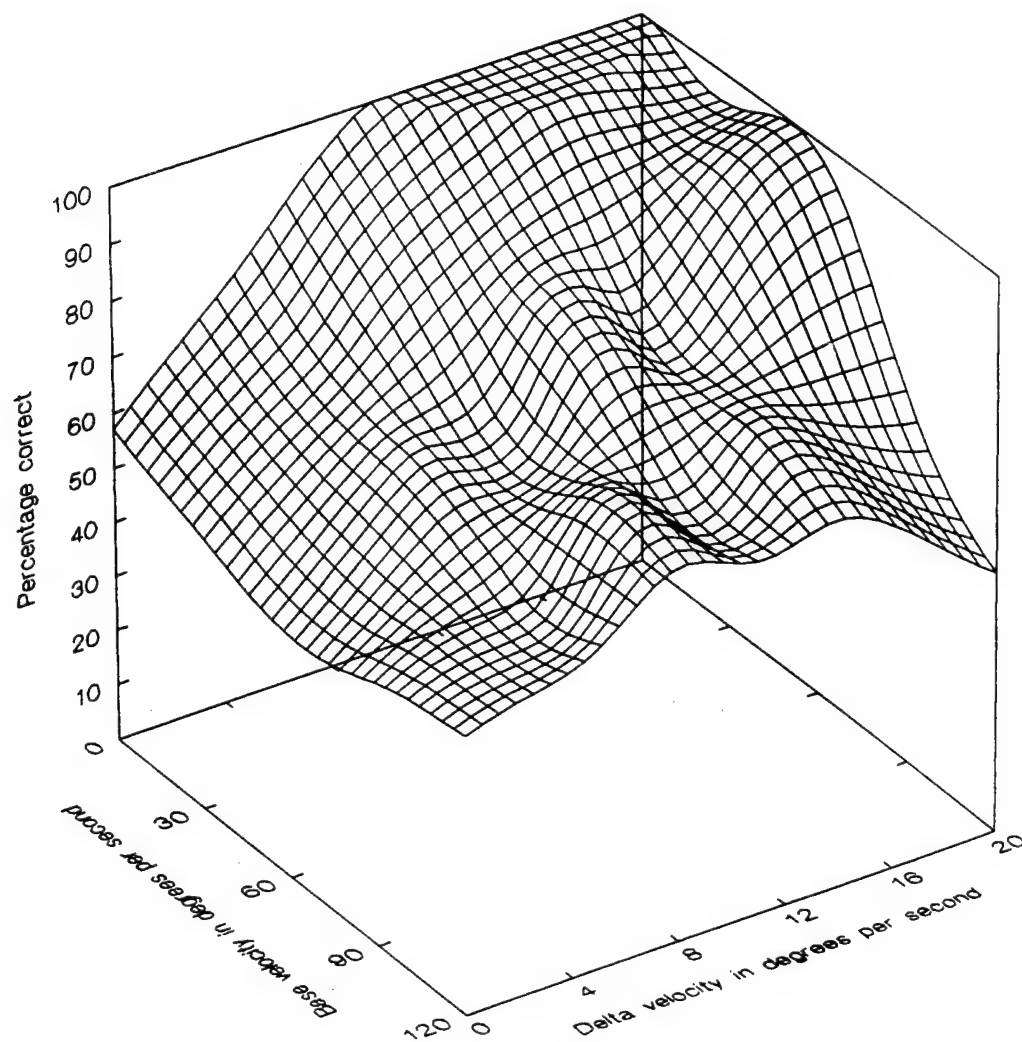


Figure 6.9: The empirical data depicted in Figure 6.8 were smoothed to form a mesh plot.

source speed. The subject was then asked to adjust a potentiometer, which controlled the speed of the moving sound source, so that the moving sound source was traveling at the individualized mile-per-hour speed associated with  $100^\circ \text{sec}^{-1}$ . The subject was given three minutes to match this speed with the moving sound source.

Perrott measured the mean and standard deviation of the moving sound source speed at the end of each 30 second time block. Perrott then utilized 67% of the standard deviation of the moving sound source speed as the measure for the Weber fraction, ignoring the initial 30 second time block. The average standard deviation of the last 5 time blocks was  $7.9^\circ \text{sec}^{-1}$  yielding a  $DL$  of  $5.3^\circ \text{sec}^{-1}$  measured at  $100^\circ \text{sec}^{-1}$  [92]. The mean of the moving sound source speed at the end of the initial 30 second time block was approximately  $101^\circ \text{sec}^{-1}$ , which was similar to the means of the other time blocks [92]. However, the standard deviation in the initial time block was approximately  $22^\circ \text{sec}^{-1}$  [92]. If the  $DL$  was computed based on this standard deviation, it would be approximately  $15^\circ \text{sec}^{-1}$  measured at  $100^\circ \text{sec}^{-1}$ , or 15%. Based on Perrott's data, it took the subjects 60 seconds of adjustment to provide data supporting the 5% Weber fraction computation.

The method used to compute the 5% Weber fraction in Perrott's experiment was not the same as the method used in this experiment, which used the method of constant stimuli. In addition, the subjects in this experiment were exposed to short duration auditory source movements as compared to those used by Perrott. Based on the methodology utilized in Perrott's study, it is reasonable to compare the resultant Weber fraction obtained in this experiment with the Weber fraction associated with the initial 30 second time-block of Perrott's data, which provides a Weber fraction estimate of approximately 15%.

In summary, the discrimination fraction of 14.1%, obtained under the presentation characteristics utilized within this experiment, indicates that the auditory localizer appears to present an auditory stimulus that adequately resembles an actual sound source for velocity discrimination tasks. It creates an ogive-shaped curve of percentage correct plotted against delta velocity, as seen in Figure 6.7, and elicits a Weber fraction for velocity discrimination which corresponds well with comparable Weber fractions documented in the literature.

Although the velocity discrimination elicited by this particular auditory localizer falls within bounds of velocity discrimination documented in the literature, the auditory presentation characteristics utilized in this study were not optimal and may have influenced the resultant Weber fraction. For example, head positioning has been found to be a powerful cue aiding auditory localization [83]. By not stabilizing the sound source location to the external environment, the human subject could not have utilized this technique and this, in turn, may have influenced the discrimination of velocity. The spectral characteristics of the audio signal utilized, a band-limited noise signal, may have also impacted localization accuracy [83] and thus, in turn, velocity discrimination. The algorithms utilized within the auditory localizer may have also contributed to influenced velocity discrimination Weber fractions by not modeling individual differences in pinna or head-form. For these reasons, it appears



prudent to evaluate the quality of the velocity perception elicited from the auditory localizer using not only velocity discrimination but also a metric of dynamic resolution. This evaluation is detailed in the following experiment.

### 6.3 Experiment three: Minimum audible movement angle

The angular extent required to distinguish moving from stationary auditory sources, which was described by Perrott as the minimum audible movement angle (MAMA), is investigated in this experiment. Specifically, this experiment evaluates the MAMA elicited using the virtual auditory localizer used within this research program. A comparison of the MAMA elicited using this virtual auditory localizer with the MAMA Perrott elicited using a real moving source provides a metric for determining dynamic equivalence between the auditory localizer and a real moving sound source.

Perrott had found the MAMA to be a function of auditory source speed [93]. He found it took approximately  $8.3^\circ$  at  $90^\circ/\text{sec}$  [93]. Perrott also stated that presentation times of at least 300 milliseconds must be used to ensure velocity perception [34]. Perrott and Tucker investigated MAMA as a function of auditory source velocity and frequency content [94]. In addition, they consolidated empirical data regarding source velocity effects on the MAMA from other researchers. This consolidation also indicated that at approximately  $90^\circ/\text{sec}$ , the MAMA, obtained from both low frequency and high frequency tones, was approximately  $8^\circ$  [94].

#### 6.3.1 Aims

The purpose of this experiment was to establish the relative quality of the motion perception elicited from this particular auditory localizer relative to a real moving sound source. The experimental design, variables, and methodology were a replica of an experiment reported by Perrott and Musicant [93] that determined threshold MAMA for a real moving sound source at several velocities. By comparing the experimental results with Perrott's findings, a correspondence may be made between perception of motion from a real sound source and from this particular auditory localizer. On the basis of the findings from experiment two, that a reasonable velocity discrimination was obtained using the localizer, it was hypothesized that a MAMA equivalent to that found using non-virtual auditory displays would be obtained.

#### 6.3.2 Independent variables

The duration of the auditory stimulus in each trial was either 50ms, 100ms, 150ms, or 300ms. In addition, the stimulus was either stationary or moving at  $90^\circ/\text{sec}$  within each trial.

### 6.3.3 Dependent variable

The experimental design incorporated a two-alternative forced-choice for the subject. In each trial, the subject decided if the stimulus was moving or stationary. The single dependent variable was the percentage decisions that were correct of the trials in which the velocity was  $90^\circ \text{sec}^{-1}$ . This variable was labeled as  $P_c$ .

### 6.3.4 Other Conditions

There were twice as many trials for each subject using stationary stimuli as moving stimuli. The stationary stimuli were presented at two horizontal locations, either straight ahead,  $0^\circ$ , or at  $8^\circ$  to the left. The movement of the auditory stimulus always began straight ahead and moved in front of the subject to the left. The auditory stimulus used in Perrott's study was a 500Hz sinusoid [93]. Unlike Perrott's experiment, a 500Hz sinusoid and a noise signal were utilized in this experiment. This difference is discussed more fully in the Results and discussion section. The spectra of the noise signal had a peak at approximately 80Hz, decreased at approximately 6dB/octave toward lower frequencies, and decreased at approximately 3dB/octave toward the higher frequencies. In addition, at approximately 21kHz, the spectra began to decrease at approximately 25dB/octave.

The experimental apparatus used in this experiment was the Localizer/CRT subject station, shown in Figure 4.3 and described in Chapter 4.

### 6.3.5 Subjects

Seven male volunteer subjects were utilized in this experiment who were college students ranging in age from 18 to 35. Two of these seven were used in a pre-experimental equipment checkout and did not participate further in experimental data collection. The remaining five subjects did not participate in the pre-experimental equipment checkout but did participate in experimental data collection. None of the subjects reported any hearing impairments. Each subject was paid for their time as a subject.

### 6.3.6 Task

Each subject was run in one session, with each session consisting of 4 blocks of 153 trials, each block differing in the duration of the stimulus. Within a block of trials, the velocity of the stimulus was randomized. Each trial lasted approximately 5 seconds. The trials were self-paced. At the beginning of the session, a discussion of the objectives of the experiment and a brief introduction to the experimental set-up was provided to the subjects. A set of training trials was then administered to each subject with feedback given after each stimulus presentation. The training normally consisted of approximately 36 trials. The first datum collection block was then

administered. The block began with a request of the subject to align the head-tracker to the external environment. This was done by looking at the center of the monitor and pressing a button inside the subject booth. Once this was accomplished, the monitor displayed a visual alert to the subject, with the auditory stimulus presented approximately 2 seconds following the onset of the visual alert. The monitor would be blanked during the auditory presentation. Following the auditory presentation, the subject was requested to press one of two buttons, indicating whether the auditory stimulus was moving or stationary. The buttons were marked with an M for moving and an S for stationary. No feedback was given to the subjects. The next trial would begin after the M or S button was pushed.

### 6.3.7 Results and discussion

Initial subjective results of this experiment were illuminating. After setting up the experimental equipment, it was noticed that a slight crackling could be heard in the headphones buried within the localized 500Hz audio signal. This crackling could only be heard when the sound source was moving. It was likely that the crackling was a result of rapid changes in amplitude of the 500Hz tone caused by moving the auditory source in azimuth angle, effectively creating an envelope over the 500Hz tone. After identifying this effect, it became a very power cue to the subject for the discrimination between the moving and the stationary auditory stimuli. It appeared that this characteristic in the audio signal could be used to gain almost perfect discrimination at even short signal durations. If this were true, it meant that the localizer could not be used to simulated real moving sound sources. To verify the apparent effect of this audio signal characteristic, two subjects were randomly selected to receive the pure tone stimuli. As anticipated, these two subjects were able to identify the moving sound source almost perfectly down to the 50ms duration level. The pattern of the data did not result in an approximation to an ogive and made the analysis technique used by Perrott inappropriate for analysis of these data. This result showed that the localizer could not be used to simulate non-virtual moving sound sources when the sound sources were emitting a 500Hz sinusoid. This limitation probably existed for any pure tone signal.

The characteristic crackling heard in the 500Hz tone could not be heard, based on subjective evaluation, when using a broad-band noise source. This experiment was continued using a broad-band noise source. Insight into the potential effect of using a broad-band source relative to a 500Hz pure tone when measuring the MAMA is necessary to enable the results obtained in this experiment to be compared with Perrott's original findings.

Much research had been done in the past and documented in the literature concerning auditory localization resolution as a function of spectral characteristics of the sound source. This resolution had been called localization blur and minimum audible angle (MAA). Strybel stated that the MAMA was affected by frequency in a manner similar to the frequency affect on MAA [119]. Research documented in the literature had found that the MAA fell between  $1^\circ - 2^\circ$  from about 100Hz to 8000Hz

with a peak of about  $4^\circ$  at 1600Hz [11] [48].

Only a few studies had evaluated localization blur as a function of source spectral characteristics using moving sound sources. Harris performed a series of localization experiments in 1972, one of which investigated localization blur with several pure tones emanating from real moving sound sources [48]. Harris found that the MAA increased when using a  $2.5^\circ/\text{sec}$  sound source relative to the MAA found from a stationary source [48]. Harris found no effect, even at 1600Hz, of sound-source spectral characteristics on MAA [48].

Perrott also found that the MAA for moving sources, which he called MAMA, increased with velocity and followed a linearly increasing relationship [93]. Perrott provided evidence of a lack of spectral effect when he pointed out in his discussion that his data corresponded to Harris's data at different sound-source frequencies [93].

It was clear from the documented literature that using pure-tones of different frequencies would affect the MAMA if the frequencies fell in specific bands, such as a 500Hz tone and a 2000Hz tone. However, the comparison of MAMAs obtained using a pure-tone of 500Hz from those obtained using a broad-band noise source was not documented in the literature. It was expected that empirically obtained MAMAs obtained using the broad-band noise could be compared to empirically obtained MAMAs obtained using a 500Hz tone with little effect being attributable to the difference in bandwidth characteristics. On the basis of these arguments, the study was continued using a broad-band noise signal.

Five subjects were utilized to determine MAMA using the noise source. The data were analyzed in the same manner as reported by Perrott, as if the data were obtained in a classical method of constant stimuli. The raw percentage scores from the individual durations were normalized to z-scores for each subject using the mean and variance obtained across each duration for that subject. A linear regression was conducted using the z-scores and the durations, yielding a functional relationship between the two. The threshold z-score were calculated using the same mean and variance using a 75% threshold value, and this z-score was then related to a duration using the regression line. A MAMA was then calculated using this duration multiplied by the velocity it was found under.

The results obtained from this experiment using the described datum analysis technique, as well as the reported results from Perrott's study, are shown in Table 6.3. The comparable results from this study and Perrott's study, in conjunction with the spectral MAA results presented by Harris [48], indicate that the auditory localizer appears to generate auditory stimuli equivalent to those produced by a real sound source moving in an anechoic chamber in terms of dynamic resolution. This result must, however, be caveated by the premise that the auditory source driving the auditory localizer be broad-band noise.

Subject	Perrott's 500Hz MAMA	Noise MAMA
1	12.4°	12.5°
2	8.9°	10.7°
3	8.2°	6.3°
4	6.0°	5.7°
5	5.9°	5.5°
Mean	8.3°	8.1°

Table 6.3: The MAMA, a dynamic resolution measure of the human auditory system, obtained with the auditory localizer was within 3% of the MAMA documented in the literature by Perrott.

## 6.4 Summary of the auditory localizer validation investigation

The static localization capability and dynamic characteristics of the specific virtual auditory display generator, or auditory localizer, used within this research program were evaluated through a review of pertinent literature and a set of experimental studies. This evaluation was necessary to assess the generalized nature of empirical results obtained using this localizer as an auditory display. This auditory localizer must have provided moving and static auditory stimuli perceptually equivalent to conventional moving or static auditory stimuli if the empirical results were to be generalized to other auditory display conditions.

A review of the literature revealed that studies by Wightman, Erickson, and McKinley supported the conclusion that the static performance, in azimuth, of this particular localizer was perceptually equivalent to a single sound source emanating within an anechoic chamber [80], [28], [140], [141].

An original, human centered, test methodology was utilized to validate the dynamic characteristics of a virtual auditory localizer as compared to a conventional auditory display. This methodology included two experiments, one establishing the ability of the localizer to generate auditory stimuli eliciting a velocity discrimination of approximately 14%, and the second establishing the minimum auditory movement angle for the localizer, which was determined to be  $8.1^\circ$  at  $90^\circ \text{sec}^{-1}$ , and which differed from previous studies in the literature by less than 3%.

The results of experiments two and three combined with the results of the literature review documenting the static characteristics of this localizer, formed the basis for the conclusion that the auditory localizer generated auditory stimuli equivalent to real sound sources moving in an anechoic chamber in terms of dynamic resolution if a broad-band audio source was utilized.

## Chapter 7

# Spatially-linked influences

### 7.1 Experiment four: Effects of moving virtual auditory stimuli on horizontal and vertical correspondence thresholds within a visual apparent motion display

The literature review reveals that the quality and type of motion perception resulting from stroboscopic visual and auditory stimuli can be affected by the temporal and spatial characteristics of the stimulus itself. The literature review also provides ample evidence in the potential of *inter-modal* influence. It is therefore, not too large a leap to hypothesize that motion perception resulting from one modality may influence the motion perception of another modality.

This experiment investigates the potential influence of stroboscopic auditory stimuli on visual stimuli which can be perceptually organized based on the visual stimuli's spatial characteristics. Specifically, in this experiment, it is hypothesized that the spatially-driven perceptual organization of stroboscopic visual stimuli can be influenced by the presence of contemporaneous stroboscopic auditory stimuli.

The visual stimuli utilized in this experiment was a 2-frame 1-dot/2-dot stroboscopic stimuli described by Kolers in his discussion of display attraction [69]. The stroboscopic visual display is shown in Figure 5.1. This figure depicts two frames of visual stimuli which alternate. The angular difference between the single dot in frame one and the lower dot in frame two is defined as the horizontal extent. The angular difference between the single dot in frame one and the upper dot in frame two is defined as the vertical extent. The horizontal extent is defined for both the visual and auditory stimuli. However, the vertical extent is not defined for the auditory display.

Kohlers described the perceptual organization elicited from the visual stimuli as being a function of a single spatial characteristic, that characteristic being the ratio of the horizontal extent to the vertical extent [69]. In this way, the 2-frame alternating stimuli could elicit perceptions of vertical motion or horizontal

motion [69]. The elicited perceptual organization would be driven by the strength of the vertical motion percept relative to the strength of the horizontal motion percept. In the vertical motion organization, a single dot would appear to move vertically with a second dot blinking to the right of the moving dot. In the horizontal motion organization, a single dot appeared to move horizontally with a second dot blinking above the moving dot. If the horizontal extent was equal to the vertical extent, a third organization might have been elicited. The third organization would appear as one dot splitting into both a horizontal dot and a vertical dot simultaneously. These three perceptual organizations are shown in Figure 5.2.

The hypothesis of this experiment, that the spatially-driven perceptual organization of stroboscopic visual stimuli can be influenced by the presence of contemporaneous stroboscopic auditory stimuli, poses the possibility that Kohlers' construction of perceptual organization strength could be affected by the presence of horizontally moving auditory stimuli. Specifically, it is hypothesized that the horizontally-moving auditory stimuli would increase the strength of the horizontal motion percept and that the ratio of horizontal extent to vertical extent that could support the horizontal perceptual organization could be increased. Put another way, it is hypothesized that by spatially and temporally linking the auditory stroboscopic display to the horizontal orientation of the visual stimuli, the perception of horizontal orientation would dominate the perception of vertical orientation due to an increase in the strength of the horizontal motion percept.

A second hypothesis focusing this experiment was that under conditions when the horizontal separation of the moving auditory stimulus was less than the spatial threshold for auditory movement detection, no influence of the visual motion perception would be created. Using Kohler's terminology, the presentation of moving auditory information below the spatial threshold for human perception would affect the organizational strength of the visual stimuli equivalently to static auditory information.

### 7.1.1 Aims

The objective of this experiment is to determine if, and to what extent, the presentation of a moving auditory signal may affect the apparent motion perception of a stroboscopic visual display in terms of the function that spatial correspondence plays in the perception. This experiment is closely tied to experiment one in that almost the identical hypotheses are evaluated using a different apparatus. A second objective of this experiment is to compare and contrast results obtained within this experiment to results obtained in experiment one.

This experiment can also be viewed as an investigation of the structure and characteristics of the process model depicted in Figures 3.25. The process model depicted within Figure 3.25, replicated to encompass the two directions of motion being utilized in this experiment, represents apparent motion perception in the horizontal and vertical directions. The combination of the two process models is



depicted graphically in Figure 5.3. The perceptual organization is modeled as a comparison of the perceptual strength output by the two process models. The horizontal process model incorporates an auditory influence path because the stroboscopic auditory stimulus is presented contemporaneously with the horizontal visual stimulus. This experiment exercises both temporal and spatial characteristics of the process model as well as involving the perceptual organization process required to combine motion detectors.

The structure of the model depicted in Figure 5.3 supports the investigation of the hypothesis that the stroboscopic auditory stimuli will increase the strength of the perception of horizontal motion. This increase in horizontal motion strength biases the comparison of horizontal to vertical motion strength resulting in a shift of threshold. It is hypothesized that by spatially and temporally linking the auditory stroboscopic display to the horizontal orientation of the visual stimuli, the perception of horizontal motion will dominate the perception of vertical motion due to an increase in the strength of the horizontal motion percept and that the increased horizontally-orientated strength would result in increased reports of horizontal organization.

In Figure 5.3, the strength of the inter-modal horizontal apparent motion perception is represented by  $R_H$  and the strength of the vertical apparent motion perception is represented by  $R_V$ . The auditory influence on  $R_H$  is modeled in Figure 5.3 by the sum of  $H_V$  and  $R_A$ . The inter-stimulus interval is constant throughout the experiment and  $I_V$  is equal to  $I_A$ . The vertical extent,  $V_V$ , is constant throughout the experiment. The visual horizontal extent,  $E_V$ , and the auditory horizontal extent,  $E_A$ , are manipulated in the experiment.

### 7.1.2 Independent variables

There were two independent variables. The first was the vertical spatial extent,  $E_V$ , which represented the scale of the visual display. The variable  $E_V$  had two values in this experiment,  $2^\circ$  or  $5^\circ$ . The second variable was the auditory presentation mode,  $S_{mode}$ , which could take on one of two values, either stationary (S) or switching (M).

In the stationary mode ( $S_{mode} = S$ ), the auditory signal was spatially coincident with the lower left visual dot. In the switching mode ( $S_{mode} = M$ ), the auditory signal alternated between the left and right visual dot and was linked with visual presentation by switching positions coinciding temporally with the single visual dot in frame one and then the lower right visual dot in frame two. The auditory stimulus remained in either the left or right position through the duration and ISI of the visual stimulus. A visual representation of these auditory presentation locations is shown in Figure 5.1.



### 7.1.3 Dependent variable

A single dependent variable was defined for this experiment. The dependent variable was the horizontal separation distance,  $E_H$ , at which the visual perception of motion direction was reported by the subject as having switched from vertical to horizontal motion or from horizontal to vertical motion.

### 7.1.4 Other conditions

The auditory signal was presented contemporaneously with the visual stimulus. The auditory signal became audible with the onset of the visual display but did not become inaudible during the ISI. The visual pattern was not view-able during the ISI. Previous work by Strybel *et al* in auditory apparent motion had shown that auditory presentations that become inaudible during the ISI tended to destroy the illusion of motion created by the rapid changing of auditory source position [120].

### 7.1.5 Subjects

Eight subjects were used in this experiment. Six subjects were male, 2 were female, and all subjects were between the ages of 18 and 35. Each subject reported no known hearing impairments. Each of the subjects had normal or corrected-to-normal visual acuity with one of the male subjects reporting slight color-blindness. Each subject was a volunteer paid for their time as a subject.

### 7.1.6 Task

The method of limits was utilized. Each subject was run in one session that lasted approximately 1.5 hours, including a verbal introduction to the overall goals of the research facility, the equipment making up the facility, reading and signing a consent form, training in the experimental booth, and finally datum collection. Datum collection consisted of 2 sets of 16 trials each for a total of 32 trials for each subject.

A short break was given to each of the subjects between sets. The trials were self-paced and each lasted approximately 90 seconds. Each trial consisted of an ascending and descending series of presentations in which the horizontal separation between the single dot in frame one and the lower dot in frame two was modified after each presentation. The end of a series was indicated by the subject reporting a change in the direction of movement from the prior presentation.  $E_H$  was recorded at the end of each ascending or descending series based on the horizontal separation between the single dot in frame one and the lower dot in frame two. The 32 trials were arranged as eight repetitions of the 4 treatment condition combinations. The order of the treatment combinations was randomized within each repetition.

Training initiated with the subjects being instructed to listen to the audio

stimulus, which was a broad-band noise source. During this period, the audio stimulus was localized in the front, to the left, to the right, and finally behind the subject. They were then instructed to listen to a demonstration of a noise source moving smoothly horizontally in the frontal plane, both left-to-right and right-to-left, for approximately 30 seconds. A set of two training trials was then administered with feedback given to the subject. The training concentrated on familiarizing the subject with how to respond using the button response box after each stimulus presentation and demonstrations of horizontal and vertical visual apparent motion perception. The first datum collection set was then administered.

The datum collection set began with a request of the subject to align the head-tracker. This action informed the computer where the subject's head was located, in position and attitude, and allowed centering of the stimulus during presentation. Alignment was done by looking at the center of the monitor and pressing a button inside the subject booth. After alignment, the monitor displayed a centered fixation target consisting of a small circle, approximately  $0.4^\circ$  in diameter and a cross approximately  $1.2^\circ$  in vertical and horizontal extent. This remained on the monitor for approximately 0.9 seconds, and was then replaced with a uniform blue field for 0.9 seconds. The blue field was approximately  $27.5^\circ$  horizontal by  $24.7^\circ$  vertical with a luminance of  $0.5 \text{ cd/m}^2$ . The blue field was followed by a dark screen ( $0.07 \text{ cd/m}^2$ ) for 0.5 seconds. Immediately preceding the stimulus presentation, and during the stroboscopic pattern, the head position and attitude of the subject was sampled to stabilize the auditory and visual image within the experimental booth.

The stimuli were presented as alternating series of ascending and descending trials, beginning with ascending, with  $E_H$  beginning as a small horizontal separation, ranging from 40% to 70% of  $E_V$ , and slowly increasing.  $E_H$  began, for the descending trials, as a large horizontal separation, ranging from 140% to 170% of  $E_V$ , and slowly decreased during the trial. The auditory and visual stimuli were then presented using a duration of 200ms, an ISI of 50ms, and repeating the 1 dot to 2 dot pattern for 7 cycles. The dots were portrayed on the monitor as filled white squares,  $0.3^\circ$  on each side, at a luminance of  $1.7 \text{ cd/m}^2$ , and drawn against a dark background of  $0.07 \text{ cd/m}^2$ . The spatial extent of the stimulus was held constant during the presentation and was modified between the stimulus presentations by the computer in a step size of  $0.18^\circ$ . After each presentation, the subject was prompted to activate a control to indicate which direction the motion occurred. Three choices were given, either horizontal, vertical, or equal. No feedback was given to the subjects. Once the subject reported the direction of motion, the visual and auditory stimuli were removed,  $E_H$  was recorded, and the sequence was repeated for the next presentation.

### 7.1.7 Results and discussion

The threshold  $E_{HS}$  obtained from each ascending and descending series were utilized to form a  $RL E_H$  by averaging each pair of  $E_H$  from each ascending and descending series pair. The horizontal separations obtained from the ascending and descending trials were  $E_{H-asc}$  and  $E_{H-dsc}$  respectively. The  $RL E_H$ ,  $E_{H-asc}$  and  $E_{H-dsc}$  are

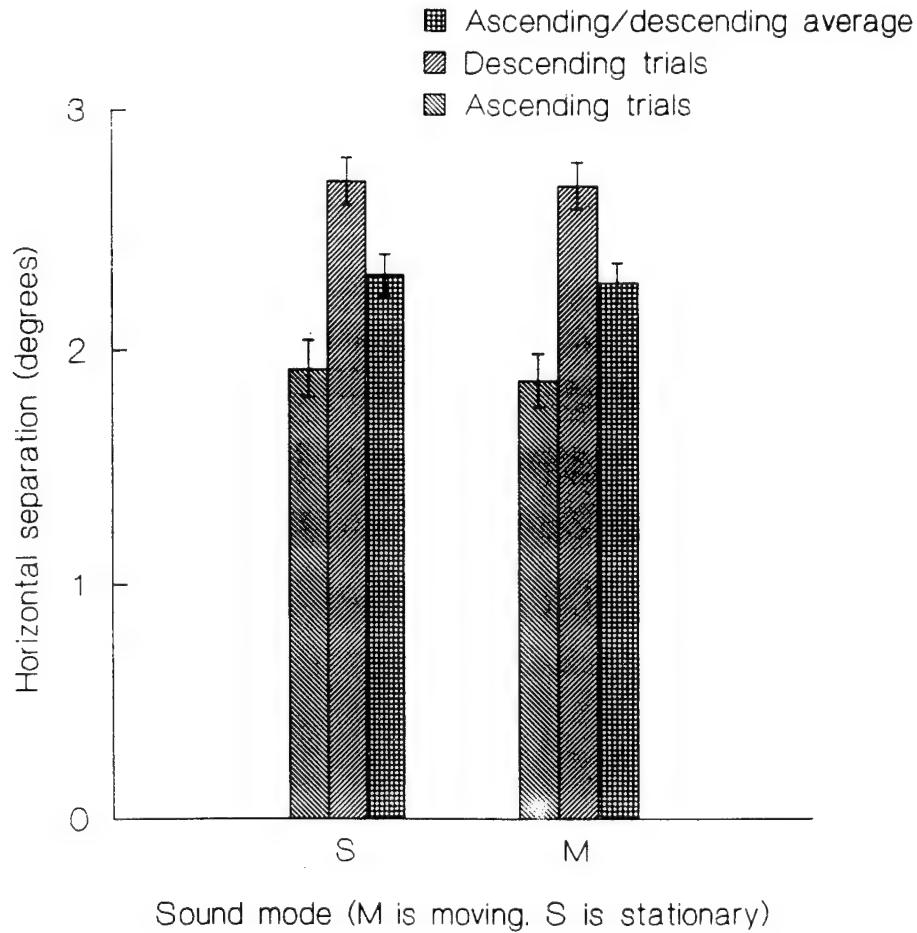


Figure 7.1: The *vertical* extent of the display,  $E_V = 2^\circ$  in this figure, significantly affected the *horizontal* extent threshold but the auditory presentation mode did not.

shown as graphs Figures 7.1 and 7.2.

Analyses of variance were performed using  $E_H$ ,  $E_{H-asc}$ , and  $E_{H-dsc}$  as affected by  $E_V$  and  $S_{mode}$ . Summaries of these separate analyses are shown in tables 7.1, 7.2, and 7.3.

The analyses of variance results indicate that  $E_V$  significantly affected  $E_H$ , [ $F(1, 7) = 728.78, p < .05$ ],  $E_{H-asc}$ , [ $F(1, 7) = 141.25, p < .05$ ], and  $E_{H-dsc}$ , [ $F(1, 7) = 337.05, p < .05$ ], which indicates that the vertical extent of the visual display significantly contributed to the absolute horizontal separation required to maintain correspondence. This is not unexpected in that the correspondence required to resolve the spatial ambiguity induced from the 1-dot to the 2-dot transition being a function of vertical spatial attributes of the visual stimuli had been documented in the literature.

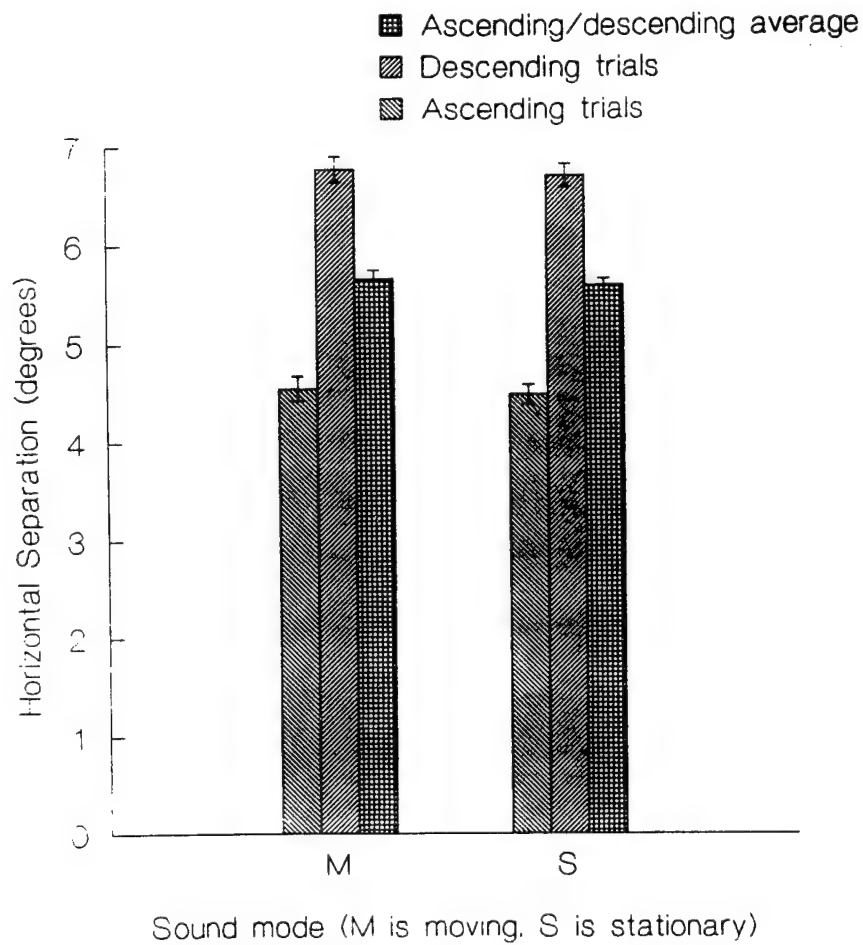


Figure 7.2: The *vertical* extent of the display,  $E_V = 5^\circ$  in this figure, significantly affected the *horizontal* extent threshold but the auditory presentation mode did not.

Source	SS	df	MS	F
$E_V$	376.81	1	376.81	728.78*
Subject	10.21	7	1.46	
Subject x $E_V$	3.62	7	0.52	
$S_{mode}$	0.24	1	0.24	1.06
Subject x $S_{mode}$	1.58	7	0.23	
$S_{mode} \times E_V$	0.48	1	0.48	1.68
Subject x $S_{mode} \times E_V$	1.99	7	0.28	

\* Significant  $p < .05$

Table 7.1: The vertical extent of the 1-dot/2-dot presentation, which was either 2° or 5°, significantly affected the averaged  $E_{HS}$ . No interaction was indicated between the horizontal extent and the auditory presentation mode.

Source	SS	df	MS	F
$E_V$	215.72	1	215.72	141.25*
Subject	28.84	7	4.12	
Subject x $E_V$	10.69	7	1.53	
$S_{mode}$	0.46	1	0.46	4.11
Subject x $S_{mode}$	0.78	7	0.11	
$S_{mode} \times E_V$	0.92	1	0.92	7.71*
Subject x $S_{mode} \times E_V$	0.84	7	0.12	

\* Significant  $p < .05$

Table 7.2: The vertical extent of the 1-dot/2-dot presentation, which was either 2° or 5°, significantly affected  $E_H$  within the ascending trials. Interaction was indicated between the horizontal extent and the auditory presentation mode.

Source	SS	df	MS	F
$E_V$	582.56	1	582.56	337.05*
Subject	21.24	7	3.03	
Subject x $E_V$	12.10	7	1.73	
$S_{mode}$	0.09	1	0.09	0.17
Subject x $S_{mode}$	3.64	7	0.52	
$S_{mode} \times E_V$	0.18	1	0.18	0.30
Subject x $S_{mode} \times E_V$	4.15	7	0.59	

\* Significant  $p < .05$

Table 7.3: The vertical extent of the 1-dot/2-dot presentation, which was either 2° or 5°, significantly affected  $E_H$  within the descending trials. No interaction was indicated between the horizontal extent and the auditory presentation mode.

The analyses of variance results also indicate that neither  $E_H$ , or  $E_{H-dsc}$  were significantly affected by the interaction of  $E_V$  and  $S_{mode}$  or  $S_{mode}$  alone. However, there is a statistically significant effect on  $E_{H-asc}$  from the interaction of  $E_V$  and  $S_{mode}$ , [ $F(1, 7) = 7.71, p < .05$ ].

The non-significant effect of  $S_{mode}$  on either  $E_H$ ,  $E_{H-asc}$ , or  $E_{H-dsc}$  indicates that the reported threshold of the visual motion perception switch from vertical to horizontal or horizontal to vertical movement is not affected by the presence of the moving auditory stimulus. The interaction of  $S_{mode}$  and  $E_V$  does not indicate a significant effect on  $E_H$  or  $E_{H-dsc}$  but does indicate a significant effect on  $E_{H-asc}$ . This provides evidence that there is a difference between the effect of the moving auditory stimulus when the display is smaller than the spatial resolution of the auditory system and the effect when the auditory stimulus is larger than the spatial resolution of the auditory system.

The graphs shown in Figures 7.1 and 7.2, as in experiment one, show a trend that is consistent with the hypothesis that the moving auditory stimulus will increase the threshold of the reported visual motion perception switching but the analysis of variance indicates that this trend is not substantiated statistically. From this statistical analysis, the hypothesis concerning the auditory influence over the visual motion perception is rejected. The second hypothesis, which concerns the lack of influence by an auditory source having spatial attributes less than required for motion detection, can not be rejected or accepted due to the rejection of the first hypothesis.

In experiment one, it was clear that some form of temporal bias was present in the data that may have masked a clear indication of results from the overall data. This type of temporal effect may have also been present within the data from this experiment. A further analysis of the data from this experiment, in a form similar to that used in experiment one, provides additional insight into the temporal aspects of the data.

The variables  $E_H$ ,  $E_{H-asc}$ , or  $E_{H-dsc}$  are shown in Figures 7.3, 7.4, and 7.5 respectively, plotted as a function of time blocks for  $E_V = 5^\circ$ . The time blocks were a temporal indicator that combined the data resulting from 4 sequential trials into a single result. In other words,  $E_H$ , when time block was 1, was the mean of  $E_H$  obtained from the first 4 trials. The second set of 4 trials was averaged for time block of 2, and so on.

These three graphs begin to show a very different look than the equivalent graphs from experiment one and also may begin to explain the lack of significant direct effect from  $S_{mode}$ . Specifically, the graphs imply that the influence of the moving auditory stimulus may not be constant over the length of the experimental session. The graphs indicate that when time block is 1, the maximum influence of the moving auditory stimulus over the visual perception occurs. A set of paired t-tests was performed to verify this indication.

The t-test was performed for each pair of datum sets at each t-block value,

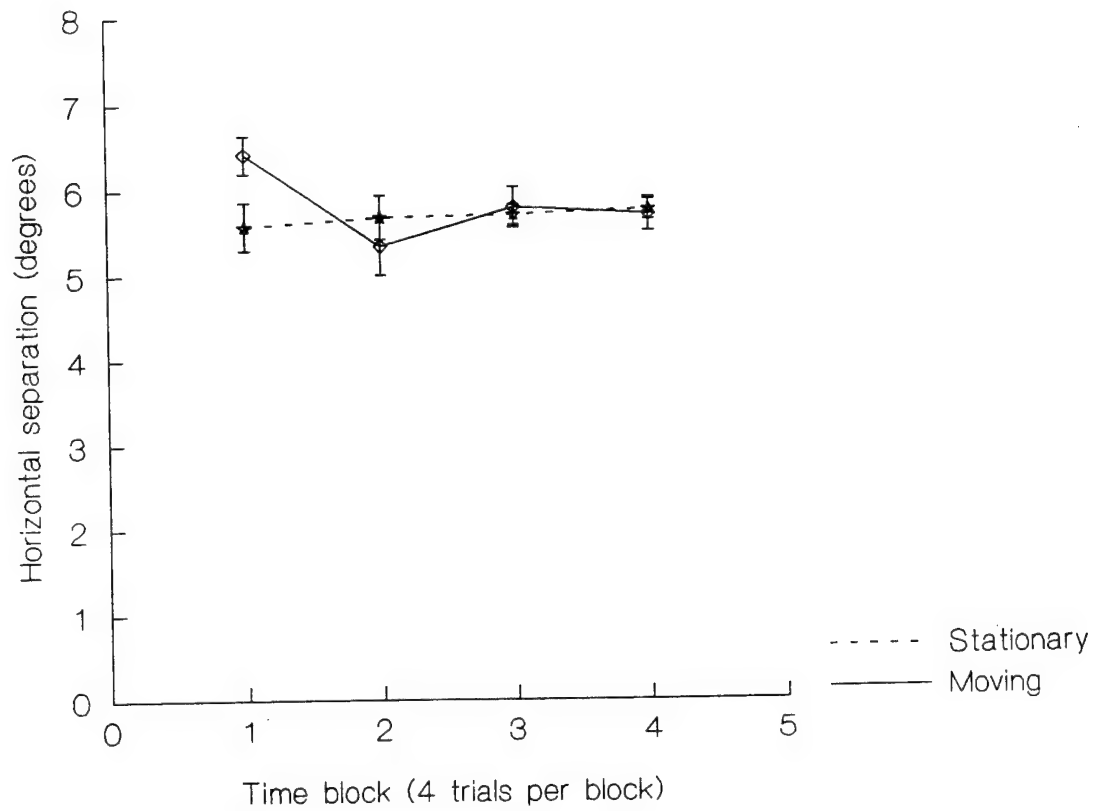


Figure 7.3: The averaged horizontal extent thresholds, when  $E_V = 5^\circ$ , were significantly shifted under the presence of the moving auditory presentation in the initial trials.

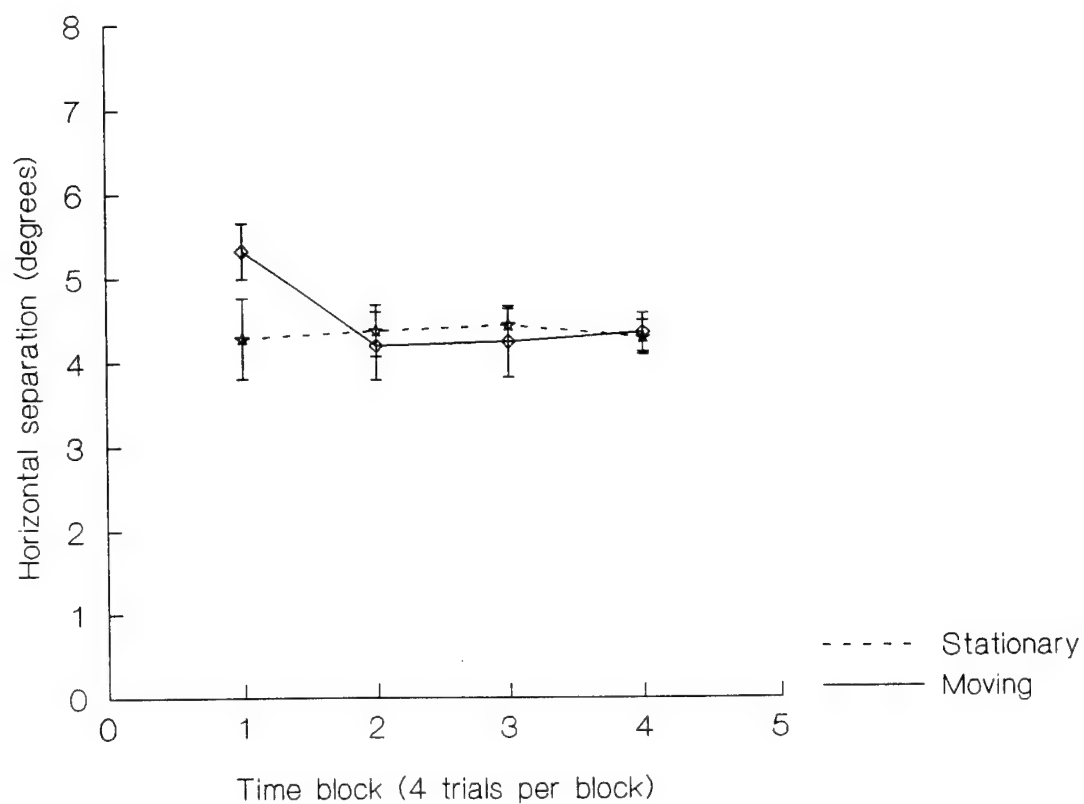


Figure 7.4: The horizontal extent thresholds from the ascending trials, when  $E_V = 5^\circ$ , were significantly shifted under the presence of the moving auditory presentation in the initial trials.



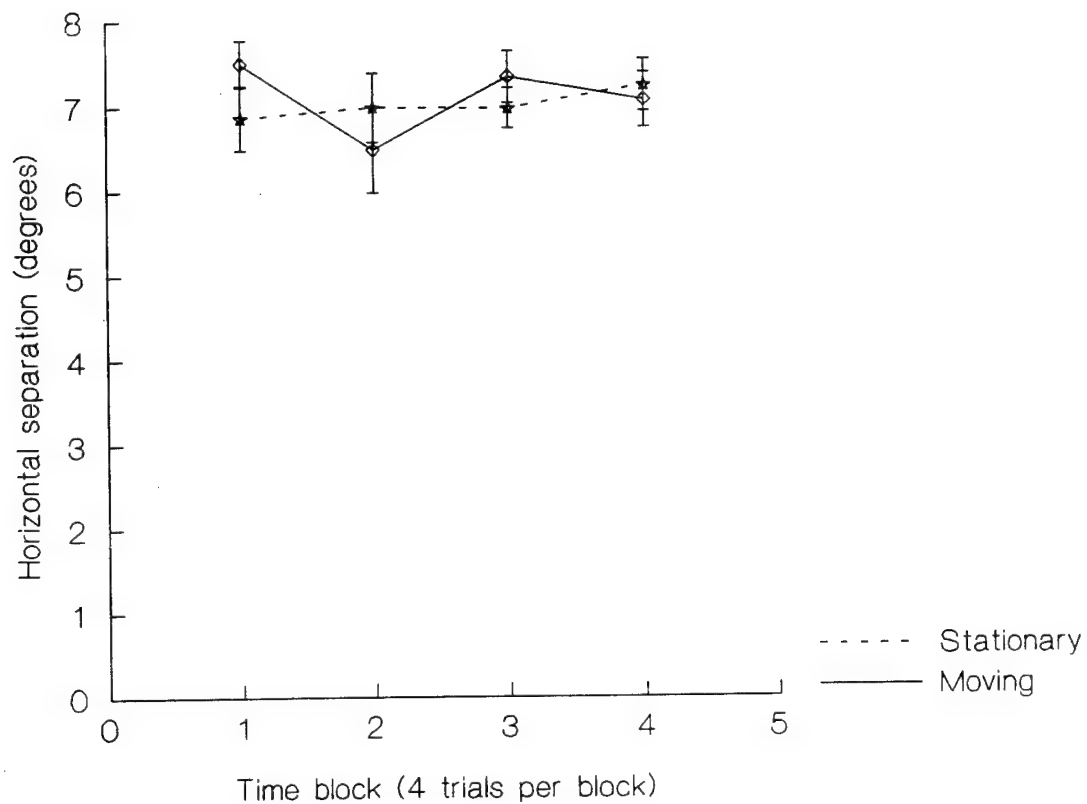


Figure 7.5: The horizontal extent thresholds from the descending trials, when  $E_V = 5^\circ$ , were not significantly shifted under the presence of the moving auditory presentation in the initial trials as they were in the ascending trials and the averaged thresholds.

checking for the possibility that the means of  $E_H$ ,  $E_{H-asc}$ , or  $E_{H-dsc}$  were different under the two conditions of  $S_{mode}$ . The results of the paired t-tests indicate that when  $E_V = 5^\circ$ , the difference of the  $E_H$  and  $E_{H-asc}$  means between the two auditory conditions are significantly different in the initial time block.

A similar analysis was performed on data obtained when  $E_V = 2^\circ$ . The variables  $E_H$ ,  $E_{H-asc}$ , or  $E_{H-dsc}$  are shown in Figures 7.6, 7.7 and 7.8 respectively, plotted as a function of time blocks for  $E_V = 2^\circ$ . It appears likely from the graphs in Figures 7.6, 7.7 and 7.8, that no statistical difference existed in thresholds obtained with the moving or stationary auditory stimulus when the vertical extent of the visual display was fixed at  $2^\circ$ . These three graphs, obtained when  $E_V = 2^\circ$ , can be compared to Figures 7.3, 7.4, and 7.5 when screening for the existence of similar temporal trends that appeared when  $E_V = 5^\circ$ .

Paired t-tests were performed on each set of time blocks for  $E_H$ ,  $E_{H-asc}$ , and  $E_{H-dsc}$  for which  $E_V = 2^\circ$ . None of the t-tests resulted in a statistically significant difference. The lack of statistical significance supports the premise that no influence exists when  $E_V$  was fixed at  $2^\circ$ .

While the auditory influence over the visual correspondence thresholds has been supported by statistical analyses under the condition of  $E_V = 5^\circ$ , the cause of the influence has not been established and several causes may be postulated.

One cause may be that the subjects attempted to provide what they thought would be the *correct* response, instead of accurately reporting only what they perceived, during the early trials in the experiment. The cause of this type of effect would be a form of experimenter-induced bias. The possibility for this to occur was minimized in this experiment through several techniques, including providing only general information to subjects regarding the hypotheses of experiments and not providing any information regarding performance that had been obtained by prior subjects or in any prior experiments. A standard written form was utilized to document the subject's consent prior to participation in the experiment that included a description of the purpose of the experiment. The consent form utilized for this experiment is shown in the appendix. In addition, a uniform training set was given to each subject before datum collection to ensure the subjects were familiar with the task. While experimenter-induced bias can not be completely ruled out in this experiment, the experimental technique reduced the probability that it may have occurred.

Another form of bias that may cause this type of effect is subject-induced bias. This type of bias would typically manifest itself in larger-than-expected differences between conditions due to the subjects' belief in the effect that should occur between conditions. However, data from this experiment do not appear to support the existence of subject induced bias. The mean thresholds shown in Figures 7.1 and 7.2 depict only slight changes in the mean thresholds from the moving to the stationary conditions. In addition, the graphs that depict temporal spread of the thresholds, show that the influence occurs only in the initial set of trials. Subject induced bias would most likely appear throughout the set of trials if it were present in these data. For these reasons, it does not appear that subject-induced bias is present in these

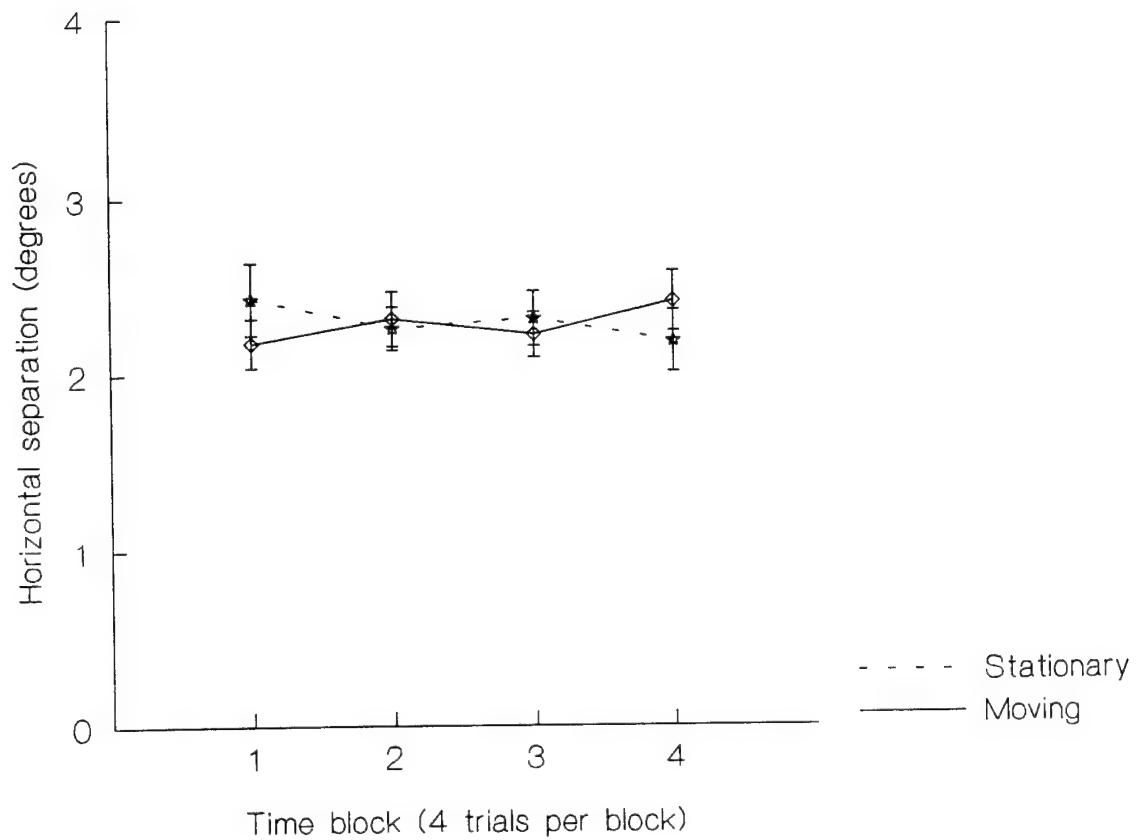


Figure 7.6: The horizontal extent thresholds from the ascending trials, the descending trials, and the average of the ascending and descending trials, when  $E_V = 2^\circ$ , were not significantly shifted under the presence of the moving auditory presentation as they were when  $E_V = 5^\circ$ . This figure depicts data averaged from the ascending and descending trials.

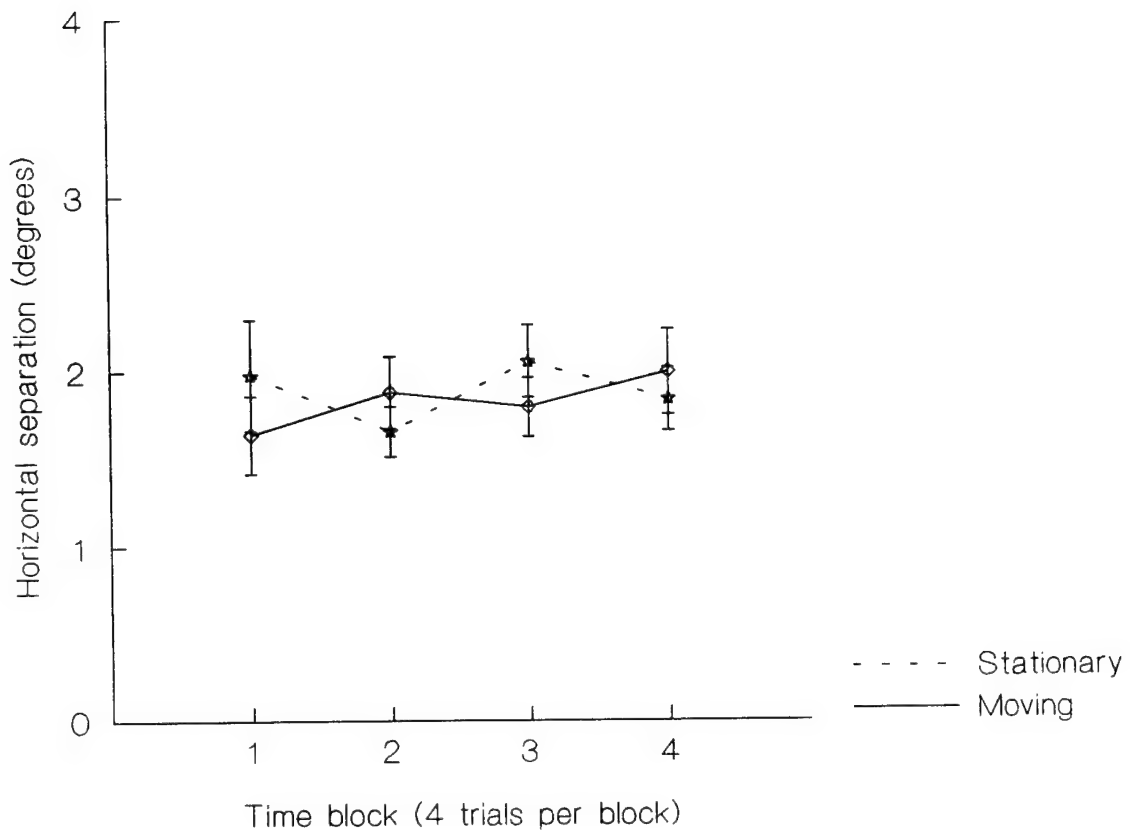


Figure 7.7: The horizontal extent thresholds from the ascending trials, the descending trials, and the average of the ascending and descending trials, when  $E_V = 2^\circ$ , were not significantly shifted under the presence of the moving auditory presentation as they were when  $E_V = 5^\circ$ . This figure depicts data from the ascending trials.

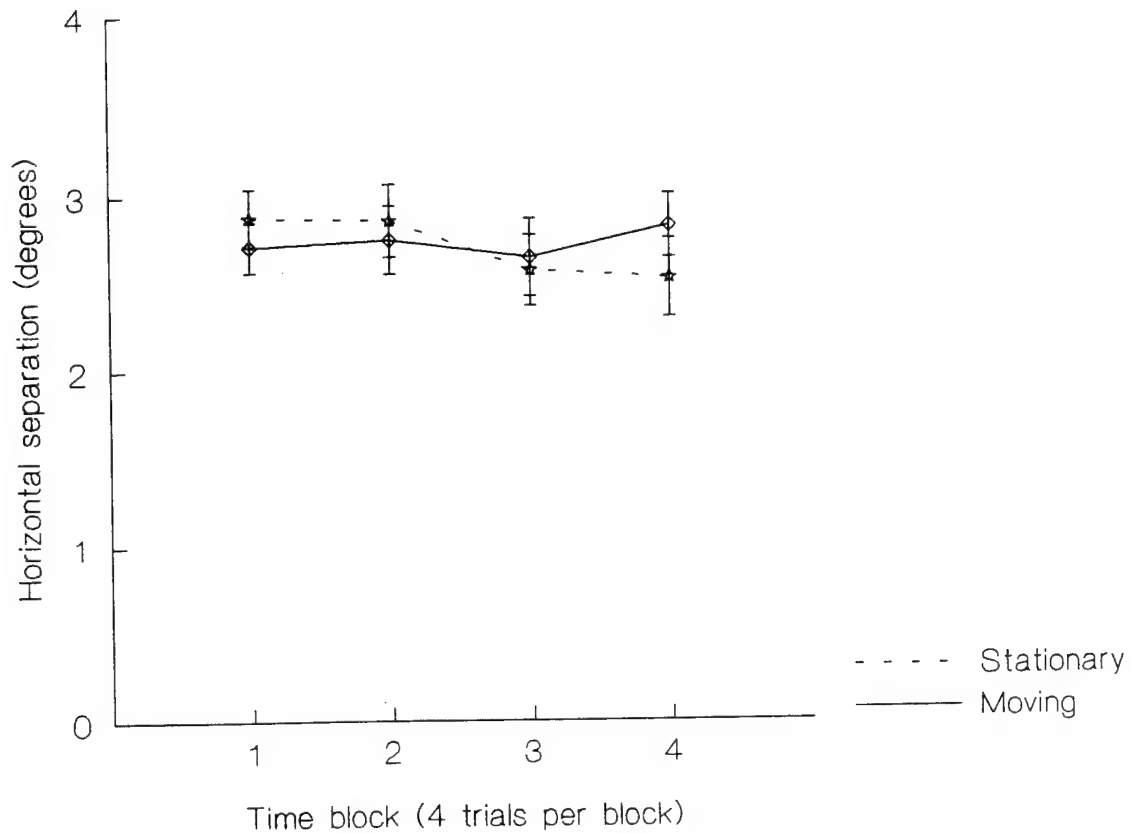


Figure 7.8: The horizontal extent thresholds from the ascending trials, the descending trials, and the average of the ascending and descending trials, when  $E_V = 2^\circ$ , were not significantly shifted under the presence of the moving auditory presentation as they were when  $E_V = 5^\circ$ . This figure depicts data averaged from the descending trials.

data.

An explanation based on the characteristics of the mechanism model may be more plausible than effects from experimenter or subject induced bias within this experiment. The ascending and descending series presents the subject with two linkages between the visual and auditory stimuli when the auditory stimulus is moving. In the case of the ascending series, the visual motion is initially perceived as being horizontal, and the auditory stimulus is also moving horizontally. As the series progresses, that linkage is reinforced until the subject reported vertical motion, at which time the visual and auditory stimuli will be in conflict. In contrast to this, the descending series begin in conflict, with strong vertical motion visually and strong auditory horizontal motion. The descending series remain in conflict until the subject reports seeing horizontal motion, at which time the auditory and visual perceptions are again unified.

As the series alternate during the experimental session the subjects may slowly de-couple the auditory and visual links to resolve the almost continual inter-sensory conflict, potentially over the course of 5 to 10 minutes. In other words, they may begin not to *listen* to the auditory stimuli. If this is the correct interpretation of these data, it would suggest that the maximum influence of the moving auditory stimulus over the visual motion perception is on the order of 15.1% (This figure is derived from the maximum change in mean from the non-moving auditory stimulus condition to the moving auditory stimulus condition in time block 1 of experiment four and can be seen in Figure 7.3). This explanation can be empirically evaluated using stimuli similar in characteristics to those used in this experiment combined with a different experimental technique.

### 7.1.8 Contrasts with experiment one

Similarities as well as dissimilarities existed between the results obtained in experiment one and in experiment four.

- 1 In experiment one, the strength of the effect, as measured by the increase in the ability to detect motion when contemporaneous auditory and visual stimuli were present, is not found in the average data nor the time-blocked data. However, a trend in the experiment one time-blocked data does appear to support the premise that a temporal spread may have been affecting the average threshold data. The subsequent analysis, using paired t-tests, does not support this premise. In experiment four, the strength of the effect, as measured by the increase in the ability to detect motion when contemporaneous auditory and visual stimuli were present, is not found in the average data but is found in the time-blocked data. The results of a paired t-test do support the premise that a temporal spread affected the average threshold data. Table 7.4 summarizes the averaged and time-blocked threshold differences between the moving and stationary auditory stimuli in experiment one and experiment four.

Experiment source	Moving	Stationary	Threshold shift
Experiment 1, averaged data	5.57°	5.35°	4.1%
Experiment 4, averaged data	5.67°	5.60°	1.2%
Maximum Block, Experiment 1 (2)	5.74°	5.41°	6.1%
Maximum Block, Experiment 4 (1)	6.41°	5.57°	15.1%

Table 7.4: Similar post-hoc statistical testing of data from experiment one and experiment four regarding the affect of time-blocked trial number indicated no effect existed in experiment one but an effect did exist in experiment four.

- 2 The variance contributed by the subjects in the within-subjects design is high in both experiments relative to the main independent variable,  $S_{mode}$ . Several potential explanations may exist for the cause of this variance. One potential explanation may be the nature of the reporting by subjects of an arbitrary threshold. The subjects' criteria may change during the course of the experiment. Hysteresis may also play a role in affecting the threshold reported by the subject. Hysteresis was found to be present by Eggleston [27] in apparent motion visual displays. In addition, hysteresis might not be constant from subject to subject.
- 3 In both experiments, the trends in the data show an influence dissipation over the course of the experiment. The dissipation is not substantiated through statistical analyses in experiment one but is substantiated in experiment four. It may, however, have been masked by other conditions within experiment one, such as the experimental method. It seems unlikely, however, that the experimental method was the cause of masking in that the same method was employed in experiment four as in experiment one because the experimental method used in experiment one was similar to that used in experiment four.

The difference in existence of the influence dissipation in experiment four, as opposed to experiment one, may be due to the fact that the auditory and visual stimuli were linked spatially and temporally in each trial in experiment four, and linked temporally but not spatially in experiment one. This difference resulted directly from the difference in display devices used in experiment one and experiment four. However, the hypothesis that the spatial linkage between the auditory and visual stimuli may have affected the influence of the auditory stimuli on the perception of the visual stimuli was not systematically manipulated as an independent variable during either of these experiments. The dissipation may have also been created by several other factors, such as fatigue or boredom on the part of the subject, as well as other display device differences.

To begin to assess the potential of continuous inter-sensory conflict being responsible for influence dissipation, it is useful to view the experimental data from

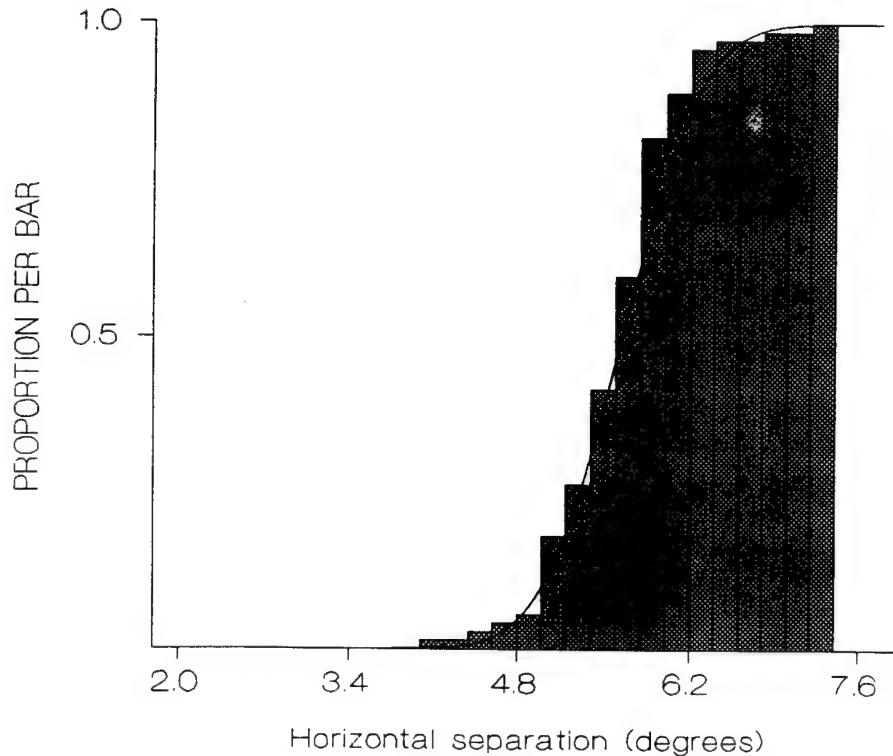


Figure 7.9: The cumulative distribution of thresholds obtained with  $S_{mode} = S$  depicted an approximate ogive curve.

experiment four in another way. By viewing the cumulative distribution of the thresholds under the moving and static conditions, it is possible to determine if the influence was evenly spread over the entire range of horizontal separations or only over a particular range of horizontal separations. In this way, it may provide a guide to the horizontal separation range useful in determining if the maximum influence can be sustained over any time period. Two cumulative distributions are shown as Figures 7.9 and 7.10.

From these distributions, it appears that the majority of the effect exists in the range of horizontal separation from  $5^\circ$  to approximately  $6^\circ$ . This range of separations can form the basis of a second investigation of the auditory influence over visual motion perception. An experimental method that may reduce the potential of inter-sensory conflict and perceptual hysteresis affecting the characteristics of any inter-sensory influence was utilized in the next experiment across horizontal separations from  $5^\circ$  to  $6^\circ$ . The details of this experiment are described in the next section.



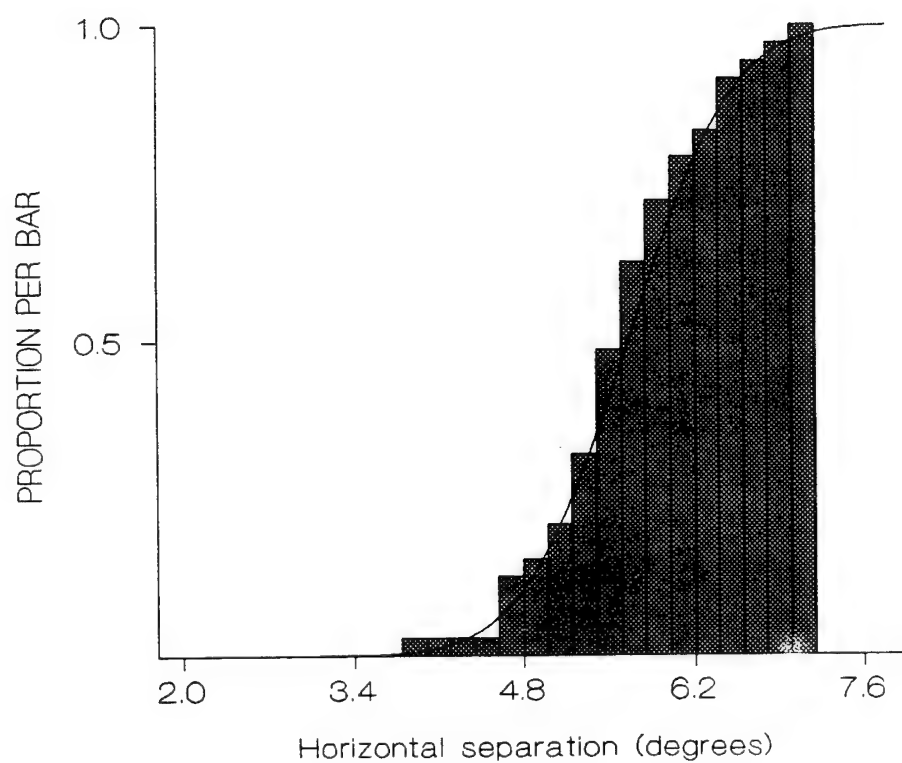


Figure 7.10: The cumulative distribution of thresholds obtained with  $S_{mode} = M$  depicted an approximate ogive curve.

## 7.2 Experiment five: An effect of auditory motion on correspondence in a visual apparent motion display, a second investigation

This experiment investigates the potential influence of stroboscopic auditory stimuli on visual stimuli which can be perceptually organized based on the visual stimuli's spatial characteristics. As in experiments one and four, it is hypothesized that the spatially-driven perceptual organization of stroboscopic visual stimuli can be influenced by the presence of contemporaneous stroboscopic auditory stimuli. Experiment five differs from experiments one and four in that it begins to address the possibility of reducing the effects of temporal decay by utilizing a different experimental technique than the method of limits utilized in experiments one and four. In addition, by concentrating datum collection around the largest threshold shifts appearing in experiment four, a more precise characterization of any potential perceptual organization strength effect may be obtained.

The visual stimuli utilized in this experiment was a 2-frame 1-dot/2-dot stroboscopic stimuli described by Kolars in his discussion of display attraction [69]. This is the same stimuli used in experiment four. The stroboscopic visual display is shown in Figure 5.1. This figure depicts two frames of visual stimuli which alternate. The angular difference between the single dot in frame one and the lower dot in frame two is defined as the horizontal extent. The angular difference between the single dot in frame one and the upper dot in frame two is defined as the vertical extent. The horizontal extent is defined for both the visual and auditory stimuli. However, the vertical extent is not defined for the auditory display.

Kohlers described the perceptual organization elicited from the visual stimuli as being a function of a single spatial characteristic, that characteristic being the ratio of the horizontal extent to the vertical extent [69]. In this way, the 2-frame alternating stimuli could elicit perceptions of vertical motion or horizontal motion [69]. The elicited perceptual organization would be driven by the strength of the vertical motion percept relative to the strength of the horizontal motion percept. In the vertical motion organization, a single dot would appear to move vertically with a second dot blinking to the right of the moving dot. In the horizontal motion organization, a single dot appeared to move horizontally with a second dot blinking above the moving dot. If the horizontal extent was equal to the vertical extent, a third organization might have been elicited. The third organization would appear as one dot splitting into both a horizontal dot and a vertical dot simultaneously. These three perceptual organizations are shown in Figure 5.2.

The hypothesis of this experiment, that the spatially-driven perceptual organization of stroboscopic visual stimuli can be influenced by the presence of contemporaneous stroboscopic auditory stimuli, poses the possibility that Kohlers' construction of perceptual organization strength could be affected by the presence of horizontally moving auditory stimuli. This is the same hypothesis which focused experiment four. Specifically, it is hypothesized that the horizontally-moving

auditory stimuli would increase the strength of the horizontal motion percept and that the ratio of horizontal extent to vertical extent that could support the horizontal perceptual organization could be increased. Put another way, it is hypothesized that by spatially and temporally linking the auditory stroboscopic display to the horizontal orientation of the visual stimuli, the perception of horizontal orientation would dominate the perception of vertical orientation due to an increase in the strength of the horizontal motion percept.

The results from experiment four indicate that the presence of moving auditory stimuli would affect the correspondence thresholds obtained when subjects viewed a 1-dot/2-dot apparent motion visual stimulus, but only in a limited way. Results from experiment four indicate that there may be a temporal decay of this effect that is attributed to the presence of conflicting inter-sensory motion cues of relatively long duration within each trial. This may have been caused by the experimental technique utilized in experiment four. In addition, results from experiment four indicate that the largest threshold shifts might occur when the horizontal separation of the display is between  $5^\circ$  and  $6^\circ$  with a vertical separation of  $5^\circ$ .

Experiment five begins to address the possibility of reducing the effects of the temporal decay by utilizing a different experimental technique than the method of limits technique utilized in experiment four. In addition, by concentrating the datum collection around the largest threshold shifts, a better characterization of any potential perceptual organization strength effect may be obtained.

### 7.2.1 Aims

The purpose of this experiment is to determine if, and to what extent, the presentation of a moving auditory signal affects the apparent motion perception of a stroboscopic visual display in terms of the function that spatial correspondence played in the motion perception organization. Two hypotheses are evaluated within this experiment. The first is that a correspondence threshold shift will exist within a  $5^\circ$  to  $6^\circ$  region of horizontal separation using the  $5^\circ$  vertical separation 1-dot/2-dot visual display utilized in experiment four. The second is that the temporal decay of the effect exhibited in experiment four will be reduced by not utilizing long exposures of potentially conflicting visual and auditory motion displays.

This experiment can be viewed similarly to experiment one and four, in that it is an investigation of the structure and characteristics of the process model depicted in Figures 3.25. The process model depicted within Figure 3.25, replicated to encompass the two directions of motion being utilized in this experiment, represents apparent motion perception in the horizontal and vertical directions. The combination of the two process models is depicted graphically in Figure 5.3. The perceptual organization is modeled as a comparison of the perceptual strength output by the two process models. The horizontal process model incorporates an auditory influence path because the stroboscopic auditory stimulus is presented contemporaneously with the horizontal visual stimulus. This experiment exercises

both temporal and spatial characteristics of the process model as well as involving the perceptual organization process required to combine motion detectors.

The structure of the model depicted in Figure 5.3 supports the investigation of the hypothesis that the stroboscopic auditory stimuli will increase the strength of the perception of horizontal motion. This increase in horizontal motion strength biases the comparison of horizontal to vertical motion strength resulting in a shift of threshold. It is hypothesized that by spatially and temporally linking the auditory stroboscopic display to the horizontal orientation of the visual stimuli, the perception of horizontal motion will dominate the perception of vertical motion due to an increase in the strength of the horizontal motion percept and that the increased horizontally-orientated strength would result in increased reports of horizontal organization.

In Figure 5.3, the strength of the inter-modal horizontal apparent motion perception is represented by  $R_H$  and the strength of the vertical apparent motion perception is represented by  $R_V$ . The auditory influence on  $R_H$  is modeled in Figure 5.3 by the sum of  $H_V$  and  $R_A$ . The inter-stimulus interval is constant throughout the experiment and  $I_V$  is equal to  $I_A$ . The vertical extent,  $V_V$ , is constant throughout the experiment. The visual horizontal extent,  $E_V$ , and the auditory horizontal extent,  $E_A$ , are manipulated in the experiment.

### 7.2.2 Independent variables

There were two independent variables. The first was the horizontal spatial extent,  $E_H$ , in which the visual and auditory display was presented. The variable  $E_H$  had four values in this experiment,  $5.04^\circ$ ,  $5.22^\circ$ ,  $5.40^\circ$ , and  $5.76^\circ$ . The second variable was the auditory presentation mode,  $S_{mode}$ , which could take on one of two values, either S or M, with S indicating a stationary auditory stimulus and M indicating a dynamic auditory stimulus.

The auditory signal was presented contemporaneously with the visual stimulus. The auditory signal became audible with the onset of the visual display and did not become inaudible during the ISI. The visual pattern was not view-able during the ISI. Previous work by Strybel *et al* in auditory apparent motion had shown that auditory presentations that become inaudible during the ISI tended to destroy the illusion of motion created by the rapid changing of auditory source position [120].

In the stationary mode ( $S_{mode} = S$ ), the auditory signal was spatially coincident with the lower left visual dot. In the switching mode ( $S_{mode} = M$ ), the auditory signal alternated between the left and right visual dot and was linked with visual presentation by switching positions coinciding temporally with the single visual dot in frame one and then the lower right visual dot in frame two. The auditory stimulus remained in either the left or right position through the duration and ISI of the visual stimulus. A visual representation of the these auditory presentation locations is shown in Figure 5.1.

### 7.2.3 Dependent variable

A single dependent variable was defined for this experiment. The dependent variable was the percentage of vertical motion reports,  $P_V$ .

### 7.2.4 Other conditions

None.

### 7.2.5 Subjects

Ten subjects were used in this experiment. Six subjects were male, 4 subjects were female, and all subjects were between the ages of 18 and 35. Each subject reported no known hearing impairments. Each of the subjects had normal or corrected-to-normal visual acuity. Each subject was a volunteer paid for their time as a subject.

### 7.2.6 Task

Each subject was run in one session that lasted approximately 1.5 hours. The method of constant stimuli was used. Each session included a verbal introduction to the overall goals of the research facility, the equipment making up the facility, the reading and signing of a consent form, training in the experimental booth, and datum collection. Datum collection consisted of 2 sets of 80 trials each for a total of 160 trials for each subject. A short break was given to the subjects between sets. The trials were self-paced and each lasted approximately 13 seconds. Each trial consisted of an 8 second presentation of the visual and auditory stimulus randomly selected from the 4 horizontal extents and the two auditory presentation modes. After each presentation, the subject was asked to report if the motion was stronger in the vertical or horizontal direction. The 160 trials were arranged as twenty repetitions of the 8 treatment condition combinations. The orders of the treatment combinations were randomized within each repetition.

Training in the experimental booth initiated with the subjects being instructed to listen to the audio stimulus, which was a noise source, as it was localized in the front, to the left, to the right, and finally behind the subject. The spectra of the noise source had a peak at approximately 80Hz, decreased at approximately 6dB/octave toward lower frequencies, and decreased at approximately 3dB/octave toward the higher frequencies. In addition, at approximately 21kHz, the spectra began to decrease at approximately 25dB/octave. They were then instructed to listen to the stimulus moving smoothly horizontally in the frontal plane, both left-to-right and right-to-left, for approximately 30 seconds. A set of sixteen training trials similar to the datum trials were then administered to the subject. The training concentrated on familiarizing the subject with how to report using the button

response box after each presentation of the stimulus and demonstration of horizontal and vertical visual apparent motion perception using a  $3.5^\circ$  and a  $6.5^\circ$  horizontal extent. The first datum collection set was then administered.

Datum collection began with a request of the subject to align the head-tracker. This action informed the computer where the subject's head was located, in position and attitude, and allowed centering of the stimulus during presentation. Alignment was accomplished by looking at the center of the monitor and pressing a button inside the subject booth. After alignment, the monitor displayed a centered fixation target consisting of a small circle, approximately  $0.4^\circ$  in diameter and a cross approximately  $1.2^\circ$  in vertical and horizontal extent. This remained on the monitor for approximately 0.9 seconds, and was then replaced with a uniform blue field for 0.9 seconds. The blue field was approximately  $27.5^\circ$  horizontal by  $24.7^\circ$  vertical with a luminance of  $0.5 \text{ cd/m}^2$ . The blue field was followed by a dark screen for 0.5 seconds. Immediately preceding the stimulus presentation, and after each cycle of the stroboscopic pattern, the head position and attitude was sampled to stabilize the auditory and visual image in the experimental booth. The auditory and visual stimuli were then presented using a duration of 200ms, an ISI of 50ms, and repeating the 1 dot to 2 dot pattern for 15 cycles. The dots were portrayed on the monitor as filled white squares,  $0.3^\circ$  on each side, at a luminance of  $1.7 \text{ cd/m}^2$ , and drawn against a dark background of  $0.07 \text{ cd/m}^2$ .  $E_H$  was randomly selected prior to each presentation from either  $5.04^\circ$ ,  $5.22^\circ$ ,  $5.40^\circ$ , and  $5.76^\circ$ .

After each presentation the subject was prompted to activate a control indicating which direction the motion occurred. Two choices were given, either horizontal or vertical. No feedback was given to the subjects. Once the subject reported the direction of motion, the sequence was repeated for the next presentation.

### 7.2.7 Results and discussion

The percentage of vertical motion reports,  $P_V$ , was calculated for each subject, averaged, and graphed in Figure 7.11 as a function of horizontal extent,  $E_H$  and auditory presentation mode,  $S_{mode}$ .

An analysis of variance was performed evaluating  $P_V$  as affected by  $E_H$  and  $S_{mode}$ . A summary is shown in table 7.5. The analysis of variance indicates that  $E_H$  does not have a significant effect on  $P_V$ . However, as can be seen from Figure 7.11, the trend of these data suggests that the effect that the moving auditory stimulus had on the visual perception at a horizontal extent of  $5.04^\circ$  may be different from the effect found at horizontal extents of  $5.22^\circ$ ,  $5.40^\circ$ , and  $5.76^\circ$ .

A second analysis of variance was performed to investigate the possibility that any effect that the moving auditory stimulus may have had on the visual perception at a horizontal extent of  $5.04^\circ$  may have been different from the effect found at horizontal extents of  $5.22^\circ$ ,  $5.40^\circ$ , and  $5.76^\circ$ . In the second analysis, all data obtained from the horizontal extent of  $5.04^\circ$  were removed and an analysis of variance was performed on the remainder of the data. The data removal was acceptable in this situation

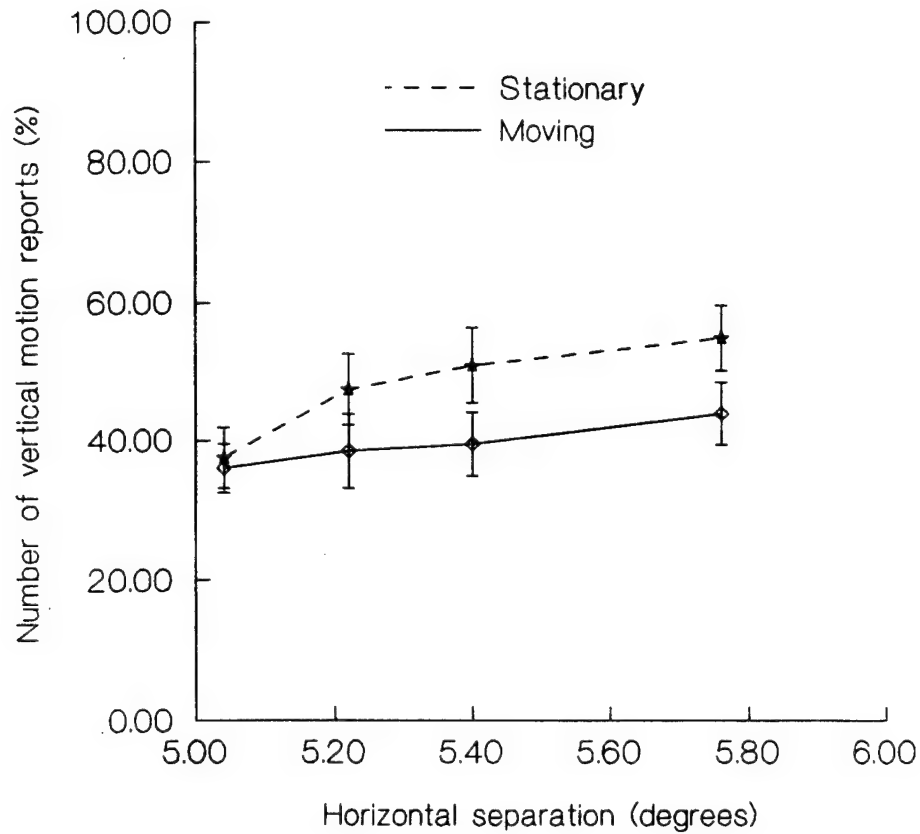


Figure 7.11: The mean percentage of vertical motion reports was significantly affected by the auditory presentation mode when the horizontal separation was 5.2° or greater.

Source	SS	df	MS	F
$E_H$	3102.5	3	1034.2	2.77
Subject	23240.0	9	2582.2	
Subject x $E_H$	10085.0	27	373.5	
$S_{mode}$	3062.5	1	3062.5	4.63
Subject x $S_{mode}$	5950.0	9	661.1	
$S_{mode} \times E_H$	772.5	3	257.5	1.19
Subject x $S_{mode} \times E_H$	5865.0	27	217.2	

Table 7.5: An analysis of variance indicated that the auditory presentation mode and the horizontal extent, which ranged from 5.02° to 5.76°, did not affect the number of vertical motion reports.



Source	SS	df	MS	F
$E_H$	761.7	2	1034.2	1.25
Subject	23700.8	9	2633.4	
Subject $\times E_H$	5471.6	18	303.9	
$S_{mode}$	3740.8	1	3740.8	5.58*
Subject $\times S_{mode}$	6034.2	9	670.5	
$S_{mode} \times E_H$	71.7	2	35.8	0.13
Subject $\times S_{mode} \times E_H$	4928.3	18	273.8	

\* Significant  $p < .05$

Table 7.6: When data obtained under the smallest horizontal extent was removed from the data set, the analysis of variance indicated that the auditory presentation mode did have a significant affect on the number of vertical motion reports. It also indicated that the horizontal extent did not have a significant affect on the number of vertical motion reports and that there was no interaction between the auditory presentation mode and the horizontal extent.

because the horizontal extent range-of-interest selected in this experiment was only approximated from results of experiment four. Summary results from the secondary analysis of variance are summarized in table 7.6.

This analysis reveals a significant effect of the presence of the moving auditory stimulus on the percentage of vertical motion reports, [ $F(1, 9) = 5.58, p < .05$ ] and thus, confirms the trend suggested in the graph of Figure 7.11. This provides evidence of the ability of a moving auditory stimulus to shift the spatial threshold necessary to determine motion direction from a 1-dot/2-dot apparent motion display.

In addition, two estimates of the strength of the effect can be obtained from this data. In both estimates, the means resulting from the  $5.04^\circ$  horizontal extent in Figure 7.11 must be disregarded. This is appropriate in that it has already been established statistically that there is no effect present at this horizontal extent. The first estimate compares the separation of the data means resulting from two auditory stimuli conditions, moving and static in Figure 7.11. The first estimate reveals an approximate 11% decrease in the number of vertical motion reports under the moving auditory condition relative to the static auditory condition.

A second estimate of effect strength can be made that is directly comparable to the maximum influence exhibited in the time-blocked data from experiment four. The maximum influence exhibited in experiment four was approximately 15%, which is graphically depicted in Figure 7.3 and referenced in Table 7.4. Two parallel lines can be constructed on Figure 7.11, one based on the three means under the moving auditory condition, and one based on the three means under the stationary auditory condition. The difference in horizontal separation between the moving auditory condition line and the stationary auditory line, found by cutting these two lines with a horizontal line at the 50% reporting rate, represents an estimate of effect strength.



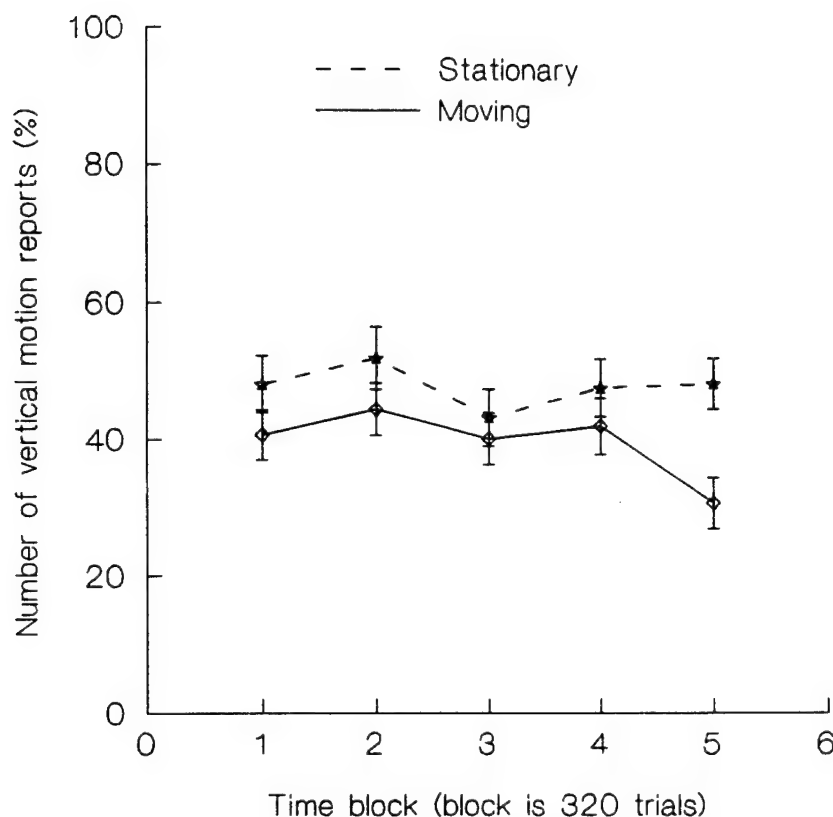


Figure 7.12: An analysis of variance performed on the percentage of vertical motion report data segmented into time blocks indicated that no significant temporal spread was present in the data.

In Figure 7.11, at a 50% rate, the stationary auditory line threshold is  $5.06^\circ$  and the moving auditory line threshold is  $6.02^\circ$ . The difference is  $0.96^\circ$  and represents a shift from the stationary auditory threshold of approximately 19%. This estimate is crude at best because only three points are used for each line estimate and the slope of each line is relatively low. However, the strength estimate of 19% compares favorably with the maximum influence exhibited in the time-blocked data from experiment four of 15%.

The temporal decay present in results from experiment one and four may also be present in the results of this experiment. To assess this possibility, data from this experiment were segmented into five time blocks, each of which represented data collected during 32 sequential trials from each subject. The percentage of vertical motion reports resulting from each subject in each block of 32 trials was calculated. The mean percentage of vertical motion reports obtained in each time block is shown graphically in Figure 7.12.

Source	SS	df	MS	F
Time block	7081.3	4	1770.3	1.65
Subject	116800.0	9	12977.8	
Subject x Time block	38606.3	36	1072.4	
$S_{mode}$	13612.5	1	13612.5	4.125
Subject x $S_{mode}$	29700.0	9	3300.0	
$S_{mode}$ x Time block	4793.8	4	1198.4	0.68
Subject x $S_{mode}$ x Time block	63143.8	36	1754.0	

Table 7.7: An analysis of variance of time-blocked data indicated that the time-block did not affect the number of vertical motion reports and did not interact with the auditory presentation mode.

An analysis of variance was performed on the segmented data to determine if the time block affected the percentage of vertical motion reports. A summary of this analysis is shown in Table 7.7. The analysis of the segmented data indicates that no significant effect resulting from time block is present. The non-significant result provides evidence supporting the premise that there is no temporal spread present within data from this experiment. The overall conclusions drawn from the spatial investigations of experiments one, four, and five, as well as the temporal investigations of experiments six and seven, are discussed in Chapter 10.

## Chapter 8

# Temporal correspondence influences

### 8.1 Experiment six: An effect of auditory motion on a temporally multi-stable visual display

Experiment one, four, and five evaluated the influence of contemporaneous auditory stimuli on *spatial* characteristics affecting visual apparent motion perception. This experiment evaluated the potential influence of contemporaneous auditory stimuli on the *temporal* characteristics affecting visual apparent motion perception. To accomplish this, a stroboscopic visual display was selected which elicited two easily-distinguishable perceptions which were affected by the temporal characteristics of the visual display.

The visual display utilized was a temporally multi-stable display composed of a two-frame three-dot stroboscopic stimulus described by Pantle [87] and sometimes referred to as a Ternus display. The stroboscopic visual display is shown in Figure 8.1.

The term multi-stable was used by Pantle to describe the alternating perceptual nature elicited when viewing stroboscopic visual stimuli [87]. The two-frame three-dot display, when viewed under certain spatial and temporal characteristics, was multi-stable using Pantle's definition [87]. Using a frame duration of 200ms and an ISI of 40ms, on the average the perception alternated 8 times per minute [87]. Pantle named the two perceptions *element* motion and *group* motion. *Group* motion occurred with ISIs  $> 40$ ms, and the visual perception was that of three dots alternating in two positions and moving as a group to the left and right [87]. *Element* motion occurred with ISIs  $< 40$ ms, and the visual perception was that of two dots being stationary and a single dot that moved from the left of the leftmost stationary dot to the right of the right-most stationary dot, seeming to pass either in front of or behind the two stationary dots [87]. The *element* and *group* motion organizations are shown in Figure 8.1.

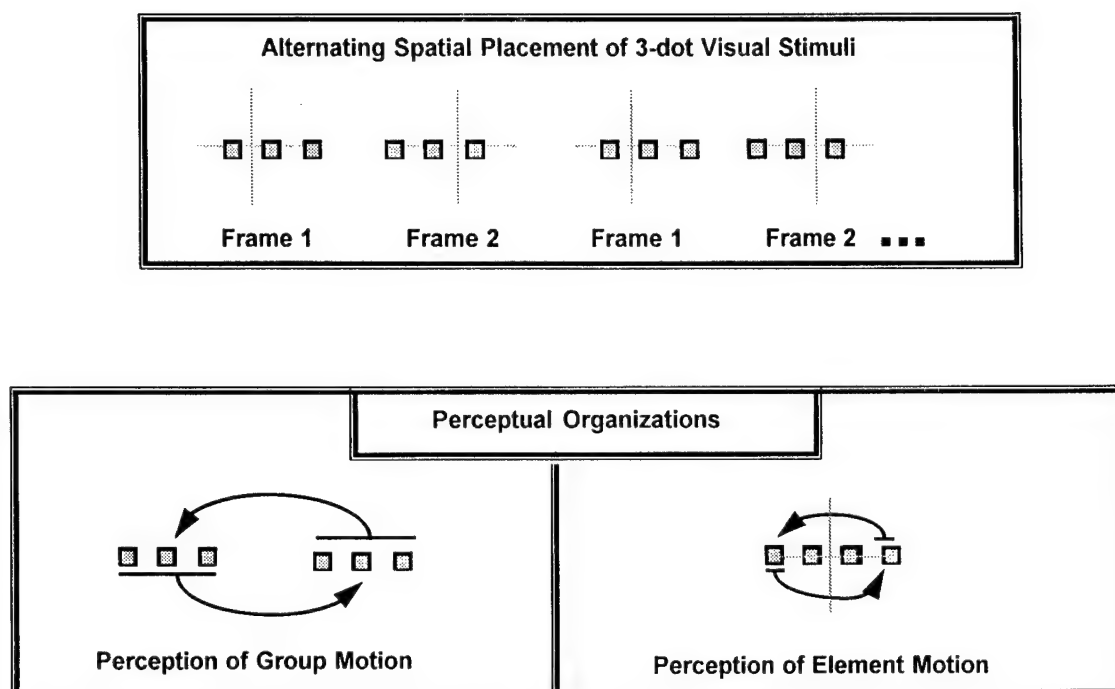


Figure 8.1: The multi-stable two-frame visual display used in experiment six elicited two distinct perceptions, *element* motion and *group* motion. The perception of *group* motion occurred at high ISIs and appeared as two groups of three dots alternating positions as a group. The perception of *element* motion occurred at lower ISIs and appeared as two stationary dots with a third dot alternating, as a single element, from the left to the right side of the stationary dots.

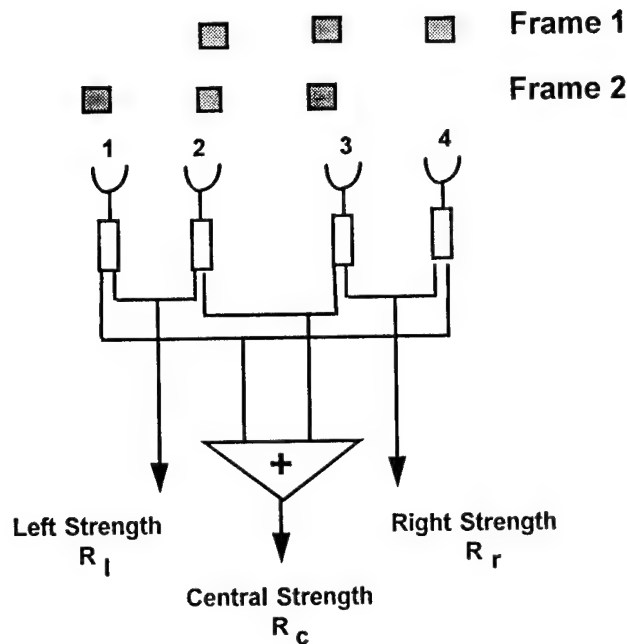


Figure 8.2: Mechanism model supporting perception of the 3-dot/3-dot visual stimulus

### 8.1.1 Aims

The purpose of this experiment is to determine if, and to what extent, the contemporaneous presentation of a moving auditory stimulus can affect the perceptual organization of a temporally multi-stable visual display. Specifically, the objective of this experiment is to evaluate if the temporal threshold, the threshold at which the *element* motion perception and the *group* motion perception would switch, can be affected by the presence of a linked auditory stimulus.

This experiment can also be viewed as an investigation of the structure and characteristics of the process model depicted in Figures 3.25. Experiment one, four, and five systematically exercised the extent processor and the decision processor depicted in the process model of Figure 3.25. However, the ISI processor, while involved, was not systematically exercised. In addition, the potential for auditory influence on the temporal integrator,  $F_t$ , was not assessed in experiment one, four, and five. This experiment assesses the potential auditory influence on the ISI processor, the decision processor, and the temporal integrator by using a visual stimulus which elicits perceptual characteristics that are functions of ISI.

A modification to the mechanism model depicted in Figure 3.22 was constructed representing the underlying mechanisms driving the perception of group motion or element motion. The mechanism model was modified such that four receptor locations were combined into three motion detectors. Figure 8.2 shows corresponding

receptor locations of the modified mechanism model and dot positions of the visual stimulus during the alternating visual frames. In Figure 8.2, the three motion perception strengths,  $R_l$ ,  $R_r$ , and  $R_c$ , are the outputs of the three motion detectors located to the left, the right, and centered on the visual stimulus. The left detector reacts to stimulus inputs between positions 1 and 2. The right detector reacts to stimulus inputs between positions 3 and 4. The central detector reacts to the sum of stimulus inputs between positions 2 and 3, and between the combined positions of 1+2 and 3+4.

The physiological basis of the bi-stable perception of group and element motion is described as being the product of inhibitory competition between parallel neural structures [87]. It is described within the literature that the perception of element motion is a result of perceived stationarity when the visual stimulus ISI is small [86] [16]. In the mechanism model, this stationarity would occur initially at receptor locations 2 and 3, shown in Figure 8.2, because the ISI between positions 1 and 4 is half of the ISI between positions 2 and 3. Grossberg linked perceived stationarity within 3-dot/3-dot stimuli to visual persistence by linking the dependence of element motion perception to stimuli conditions such as background luminance, stimulus contrast, dot size, and stimulus duration [45]. Visual persistence caused by temporal filtering within the motion perception system is well documented within the literature [133] [25] and included in many of the models of motion perception discussed in the literature review of this report.

The effect of perceived stationarity on the perception of group or element motion can be modeled based upon a competition of the center motion detector and the left and right motion detectors shown in Figure 8.2. When the ISI of the visual stimulus is large eliciting a perception of group motion, the magnitudes of the three perceptual strengths are approximately equal. As the ISI of the visual stimulus decreases, the visual persistence, and the corresponding perceived stationarity, at stimulus positions 2 and 3 increases. The visual persistence affects the detector by decreasing the resultant output. As the output of the central detector decreases, the prominence of the left and right detectors begins to dominate the input to the neural competition increasing the likelihood of perceiving element motion relative to perceiving group motion. The output of the portion of the central detector responding to visual stimulus inputs between position 2 and 3 is not a function of ISI and is always zero.

The process model shown in Figure 10.1, which is only a slight derivative of Figure 3.25, supports the assessment of a potential influence of moving auditory stimuli on visual apparent motion perception. The process model is abstracted from the mechanism model shown as Figure 8.2. The process model depicted in Figure 10.1 includes three motion-detector outputs driving the decision processor, which includes the neural competition. The output of the decision process is a report of either group motion or element motion. The process model depicted in Figure 10.1 includes the potential inter-model influence to be assessed within this experiment. In addition, the methodology used in this experiment does not inhibit the effect of the temporal integrator on visual apparent motion perception and thus, enables an initial assessment of the potential for auditory influence on visual

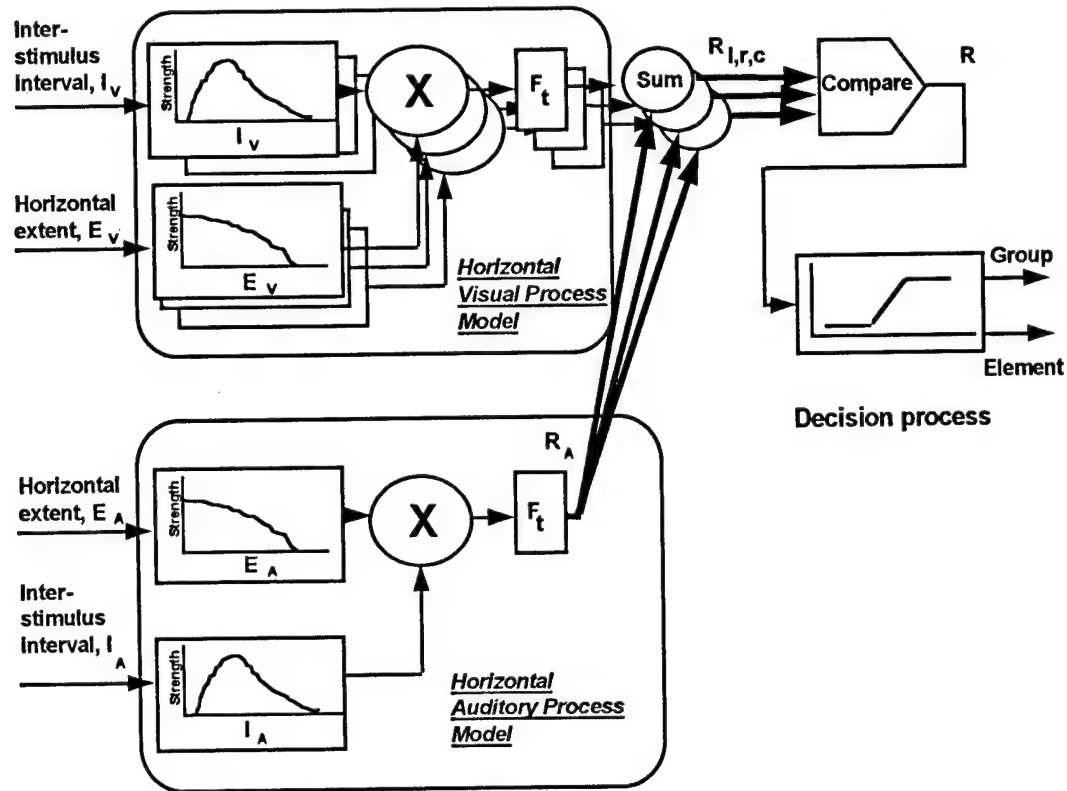


Figure 8.3: Process model supporting perception of the 3-dot/3-dot visual stimulus

apparent motion perception with hysteresis present to be accomplished.

It could be argued intuitively that the auditory stimulus would enhance the ability of the subject to capture and maintain the *group* motion perception. This effect would then manifest itself as a lowering of the ISI at which a switch from *element* motion perception to *group* motion perception or *group* motion perception to *element* motion perception will occur. However, the mechanism and process models in Figure 8.2 and Figure 10.1 do not support this position. Because the position of the auditory stimuli moves between receptor position 1 and 4, the influence of the auditory stimuli would equally affect each of the three motion-detector outputs. In this way, the magnitude of each of the detectors outputs may be affected but the result of the comparison process would not be affected. Thus, the hypothesis of this experiment is that the presence of the moving auditory stimuli will not affect the ISI at which a switch from *element* motion perception to *group* motion perception, or *group* motion perception to *element* motion perception, will occur.

### 8.1.2 Independent variables

There were two independent variables. The first was the spatial extent,  $E_H$ , in which the visual display was presented. The variable  $E_H$  had two values in this experiment,  $3.0^\circ$  or  $9.0^\circ$ .

The second variable was the auditory presentation mode,  $S_{mode}$ , which could take on one of three conditions, either stationary, switching, or smooth. Of these three conditions, both the switching and smooth values represented moving auditory presentations. The stationary condition represented the non-moving auditory stimulus presentation.

In the stationary mode, the auditory signal was presented as a localized source in front of the subject. In the switching mode, the auditory signal was linked with visual presentation by switching positions coinciding spatially with the left-most dot in the left frame and the right-most during the right frame and remaining in that position through the duration and ISI. In the smooth mode, the auditory source remained fixed during the duration time, but smoothly moved to the next position during the ISI such that it would coincide spatially with the position of the outside dot at the beginning of the duration. The three auditory presentation modes are shown in Figure 8.4.

### 8.1.3 Dependent variable

A single dependent variable was defined for this experiment. The dependent variable was the ISI at which the visual perception changed from *group* motion to *element* motion or from *element* motion to *group* motion. This variable was labeled  $T_{ISI}$  and represented, for each trial, the threshold ISI.

### 8.1.4 Other conditions

The auditory signal was presented contemporaneously with the visual three-dot stimulus. The auditory signal became audible with the onset of the visual display but did not become inaudible during the ISI. The three-dot visual pattern was not view-able during the ISI. The audio signal was a noise signal. The spectra of the noise signal had a peak at approximately 80Hz, decreased at approximately 6dB/octave toward lower frequencies, and decreased at approximately 3dB/octave toward the higher frequencies. In addition, at approximately 21kHz, the spectra began to decrease at approximately 25dB/octave.

The auditory signal was presented contemporaneously with the visual stimulus. The auditory signal became audible with the onset of the visual display and did not become inaudible during the ISI. The visual pattern was not view-able during the ISI. Previous work by Strybel *et al* in auditory apparent motion had shown that auditory presentations that become inaudible during the ISI tended to destroy the illusion of motion created by the rapid changing of auditory source position [120].

The duration of the visual stimulus during the on periods was 100ms. The ambient sound pressure level within the test booth was 48 dB(A). The sound pressure level within each ear-cup of the headset, when the auditory localizer was positioned at 0° relative to the subject, was 75 dB(A). The attack envelope of the auditory stimulus



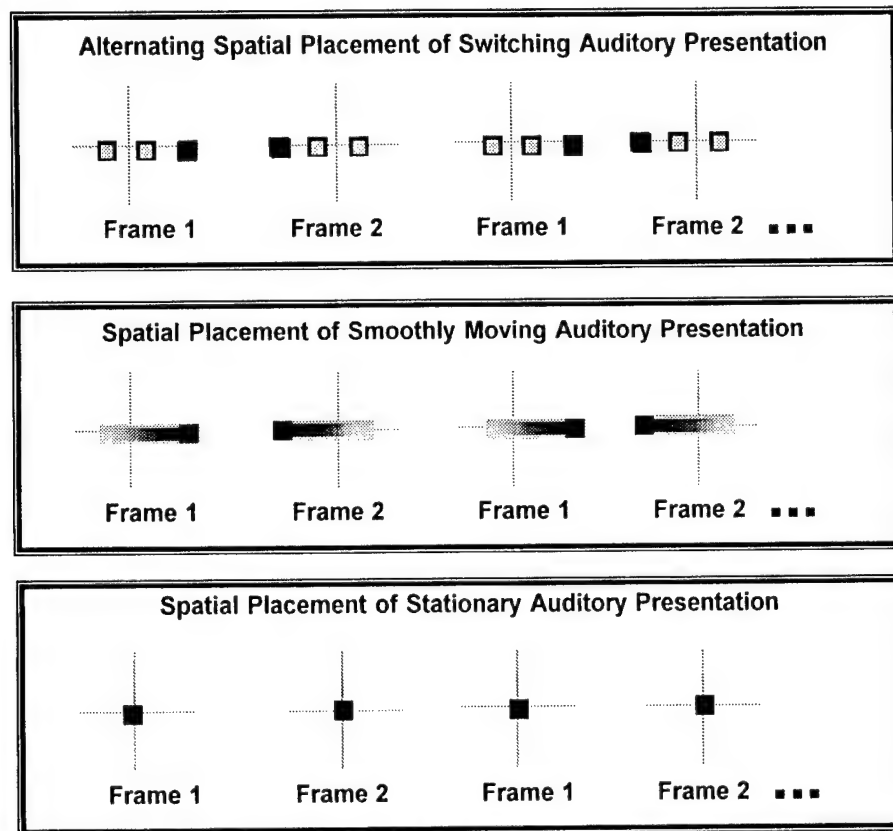


Figure 8.4: Three auditory presentation modes were used in experiment six. In the stationary mode, the auditory signal was presented as a localized source in front of the subject. In the switching and smooth modes, the auditory signal was linked with visual presentation by switching positions coinciding spatially with the left-most dot in the left frame and the right-most during the right frame.

was 60ms. The decay envelope of the auditory stimulus was 30ms. The background luminance of the CRT within the booth was  $0.074 \text{ cd/m}^2$  (0.022 Ft-L). The luminance of each dot portrayed on the CRT was  $2.90 \text{ cd/m}^2$  (0.78 Ft-L). The spatial extent of the visual stimulus was either  $9.0^\circ$  or  $3.0^\circ$  as measured from the center of the leftmost dot to the center of the rightmost dot. Each visual dot was square and was  $0.1^\circ$  on each edge.

### 8.1.5 Subjects

Eleven subjects were used in this experiment. Seven subjects were male, 4 subjects were female, and all subjects were between the ages of 18 and 35. Each subject reported no known hearing impairments. Each of the subjects had normal or corrected-to-normal visual acuity with one of the male subjects reporting slight color-blindness. Each subject was a volunteer paid for their time as a subject.

### 8.1.6 Task

The method of limits was utilized. Each subject was run in one session, with each session consisting of 2 blocks of 48 trials each. Within each block of trials, combinations of extent and auditory presentation were group-randomized, such that each of the six combinations of extent and auditory presentation mode was used once in a group. Each group was organized as an ascending trial followed by a descending trial, yielding 12 trials to a group with 4 groups to a block.

At the beginning of the first session for each subject, a discussion of the objectives of the experiment was presented as well as a brief introduction to the experimental set-up. Each subject read and signed a human-use consent form. A set of training trials was then administered with instructions given to each subject after each trial. The training consisted of approximately 20 trials. The first 12 trials trained the *element* and *group* motion visual perceptions and the remaining trials were used to train the detection of the perception switch and the activation of a control by the subject when the visual perceptual switch appeared to occur. The first datum collection block was then administered.

The datum collection block began with a request of the subject to align the head-tracker, which initiated the computer's measurement of the subject's head position and attitude. This initial measurement allowed centering of the stimulus presentation. To accomplish the alignment, the subject looked at the center of the monitor and pressed a button inside the subject booth.

Once alignment was accomplished, the monitor displayed a centered fixation target and a surrounding circle filled to indicate the percentage of trials remaining in the block. This remained on the monitor for approximately 3 seconds, and then was replaced with a filled screen for 1 second followed by a blank screen for 1 second. A thin border was then drawn around the edge of the blank screen of the monitor and

Source	SS	df	MS	F
$S_{mode}$	2044.22	2	1022.11	2.64
Subject	105909.91	10	10590.99	
Subject x $S_{mode}$	7730.82	20	386.54	
$E_H$	160306.52	1	160306.52	135.99*
Subject x $E_H$	11788.19	10	1178.82	
$S_{mode} \times E_H$	1627.03	2	813.51	1.96
Subject x $S_{mode} \times E_H$	8320.52	20	416.03	

\* Significant  $p < .05$

Table 8.1: During the ascending trials, only the horizontal extent of the 3-dot/3-dot display significantly affected the ISI reported at the perceptual switch from *element* to *group* motion.

remained there for 0.5 seconds. The stimulus was then presented. Immediately preceding the stimulus presentation, the head position and attitude were sampled to stabilize the auditory image in the experimental booth.

The stimuli were presented as alternating series of ascending and descending trials. The trials were self-paced and each lasted approximately 23 seconds. The ascending trials were initiated with an ISI ranging from approximately 8ms to 17ms, and then continued with increasing ISIs over the duration of the trial. The descending trials were initiated with an ISI ranging from approximately 167ms to 183ms and continued with decreasing ISIs over the duration of the trial. The ISI was modified by the computer during the stimulus presentation in a monotonic sequence. Each trial was ended by activation of the control by the subject.

The subject was instructed to activate a control when the visual perception switched, either from *element* to *group* motion on ascending trials or *group* to *element* motion on descending trials. No instruction was given to the subjects during the datum collection session. Once the subject activated the switch, the visual and auditory stimuli were removed and the fixation target was drawn for 1 second. The sequence was then restarted for the next trial. The 96 trials were given to each subject in the course of an hour. A five minute break was given to the subjects between blocks.

### 8.1.7 Results and discussion

The  $T_{ISI}$  obtained from each ascending and descending trial was utilized to form a  $RL T_{ISI}$  by averaging each  $T_{ISI}$  as a pair. The mean threshold  $T_{ISI}$  resulting from this experiment for the ascending and descending trials, as well as the averaged  $RL T_{ISI}$ , is graphically depicted in Figures 8.5 and 8.6.

An analysis of variance was performed evaluating potential main effects and interactions of the three auditory presentations and the horizontal extent on  $T_{ISI}$ .

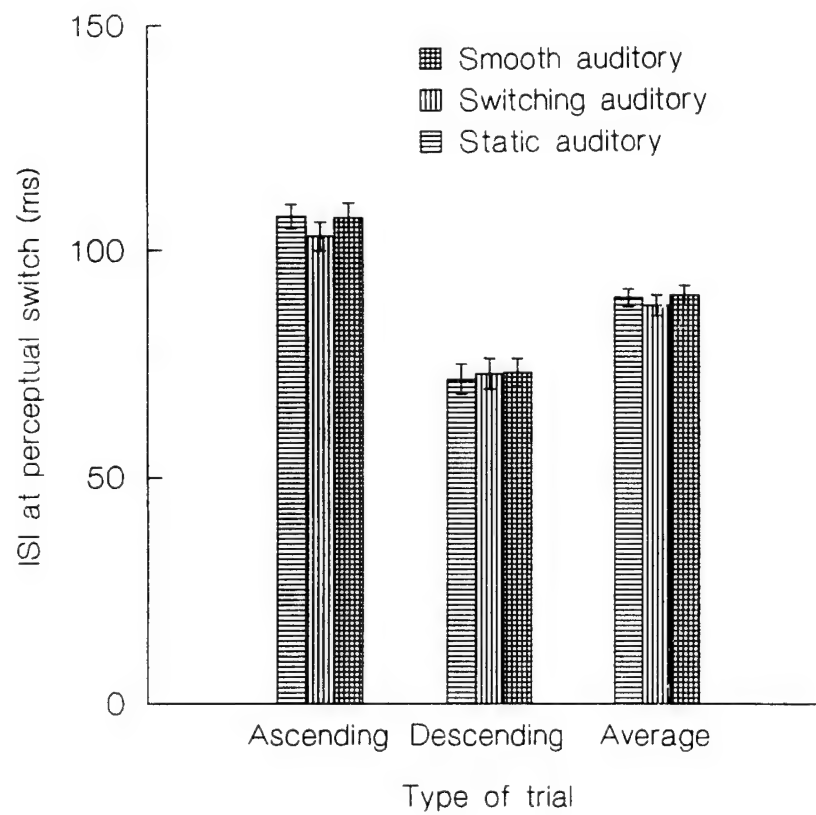


Figure 8.5: The threshold ISIs obtained at a horizontal extent of  $3^\circ$  were not significantly affected by the auditory presentation mode.

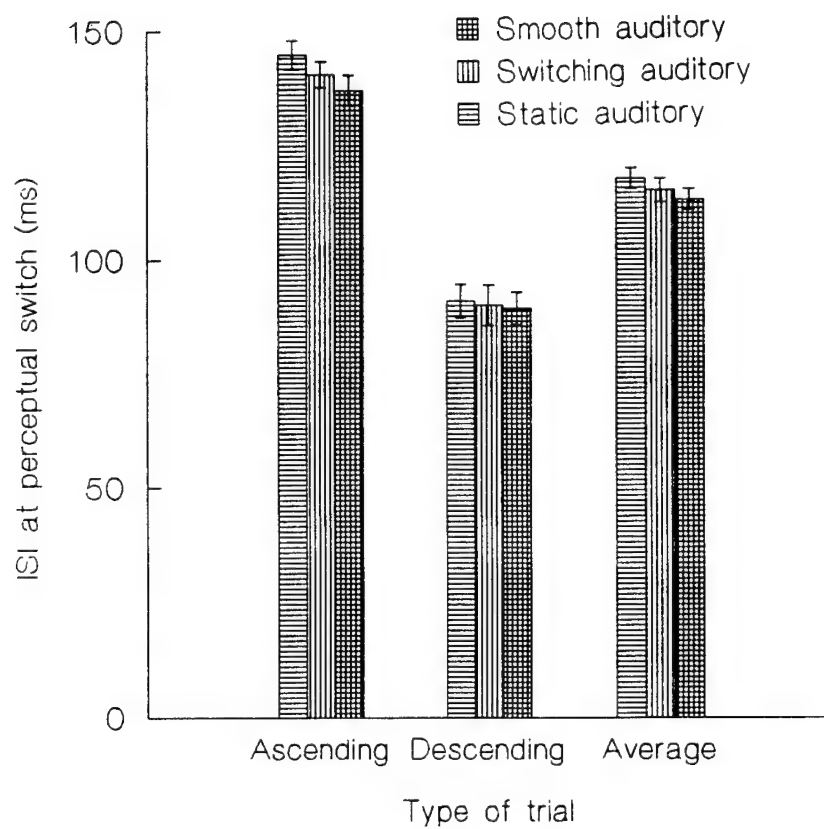


Figure 8.6: The threshold ISIs obtained at a horizontal extent of  $9^\circ$  were not significantly affected by the auditory presentation mode.

Source	SS	df	MS	F
$S_{mode}$	2.82	2	1.41	0.00**
Subject	191010.99	10	19101.10	
Subject x $S_{mode}$	18260.23	20	913.01	
$E_H$	39839.14	1	39839.14	16.86*
Subject x $E_H$	23631.28	10	2363.13	
$S_{mode}$ x $E_H$	223.02	2	111.51	0.24
Subject x $S_{mode}$ x $E_H$	9469.20	20	473.46	

\*\* Low F-value

\* Significant  $p < .05$

Table 8.2: During the descending trials, only the horizontal extent of the 3-dot/3-dot display significantly affected the ISI reported at the perceptual switch from *group* to *element* motion.

Source	SS	df	MS	F
$S_{mode}$	490.35	2	245.18	0.75
Subject	41921.94	10	4192.19	
Subject x $S_{mode}$	6579.55	20	328.98	
$E_H$	89994.12	1	89994.12	97.11*
Subject x $E_H$	9267.03	10	926.70	
$S_{mode}$ x $E_H$	677.77	2	338.89	1.08
Subject x $S_{mode}$ x $E_H$	6266.19	20	313.31	

\* Significant  $p < .05$

Table 8.3: The averaged ISI, reported at the perceptual switch from *group* to *element* or *element* to *group* motion, was significantly affected only by the horizontal extent of the 3-dot/3-dot display.

The analysis of variance is summarized for the ascending trials, the descending trials, and the averaged trials in Tables 8.1, 8.2, and 8.3 respectively.

The horizontal extent of the stimulus,  $E_H$ , causes a shift in the average  $RL$  of  $T_{ISI}$  as well as the ascending and descending  $T_{ISIs}$ , [ $F(1, 10) = 97.11, p < .05$ ], [ $F(1, 10) = 135.99, p < .05$ ], and [ $F(1, 10) = 16.86, p < .05$ ] respectively. The auditory presentation mode,  $S_{mode}$ , does not appear to contribute significantly to  $T_{ISI}$ . There is no interaction between  $E_H$  and  $S_{mode}$ .

One result of this experiment is that reduction of the stimulus spatial extent, represented by  $E_H$ , reduces the threshold ISI necessary for eliciting perceptual shifts. This can be seen graphically by comparing Figures 8.5 and 8.6. This is not unexpected in that it adheres to one of Korte's Laws relating ISI to spatial separation of an apparent motion display.

Source	SS	df	MS	F
$S'_{mode}$	2032.69	1	2032.69	6.01*
Subject	97709.37	10	9770.94	
Subject x $S'_{mode}$	3381.88	10	338.19	
$E_H$	147398.74	1	147398.74	118.00*
Subject x $E_H$	12490.98	10	1249.10	
$S'_{mode} \times E_H$	373.35	1	373.35	1.75
Subject x $S'_{mode} \times E_H$	2129.06	10	212.91	

\* Significant  $p < .05$

Table 8.4: When the two moving auditory presentation modes were combined into one, the auditory presentation mode and horizontal extent both significantly affected the ISI reported at the perceptual switch from *element* to *group* motion in the ascending trials.

Both the 3° and 9° displays incorporated spatial extents that were larger than the minimum static resolution of the human's auditory localization capability as documented in the literature. Thus, the lack of interaction between  $E_H$  and  $S_{mode}$  indicates that main effects are consistent across independent variable manipulations.

Although the mean motion perception threshold ISI appears to be lowered by the presence of auditory motion as seen in Figures 8.5 and 8.6, this is not a statistically significant difference. The conclusion from the analysis is that the presence of the auditory stimulus did not affect the visual motion perception. However, it is possible that the lack of statistical significance is caused by a lack of statistical power resulting from utilizing too few subjects, or too many conditions, or too few trials in the experiment. A secondary analysis was performed to provide some insight into these potentials.

To accomplish the secondary analysis, the three auditory presentation types were combined into two types to increase datum counts at each presentation type. Data from the two auditory presentations containing motion, those being switching motion and smooth motion, were combined into a single presentation mode. The stationary auditory presentation made up the second mode. The two new modes were represented by a new independent variable,  $S'_{mode}$ . The mean thresholds for the ascending and descending  $T_{ISI}$  and  $RL T_{ISI}$  were re-calculated and are depicted in Figures 8.7 and 8.8.

The secondary analysis of variance was conducted on the threshold and  $RL T_{ISI}$  using horizontal extent,  $E_H$ , and the re-categorized auditory presentations,  $S'_{mode}$ , using a general linear model procedure to account for the differing number of data within cells due to the re-categorization. The secondary analysis of variance is summarized in Tables 8.4, 8.5, and 8.6 relating to the ascending, descending, and average  $T_{ISI}$  respectively.

The secondary analysis, as in the primary analysis, reveals a significant effect of

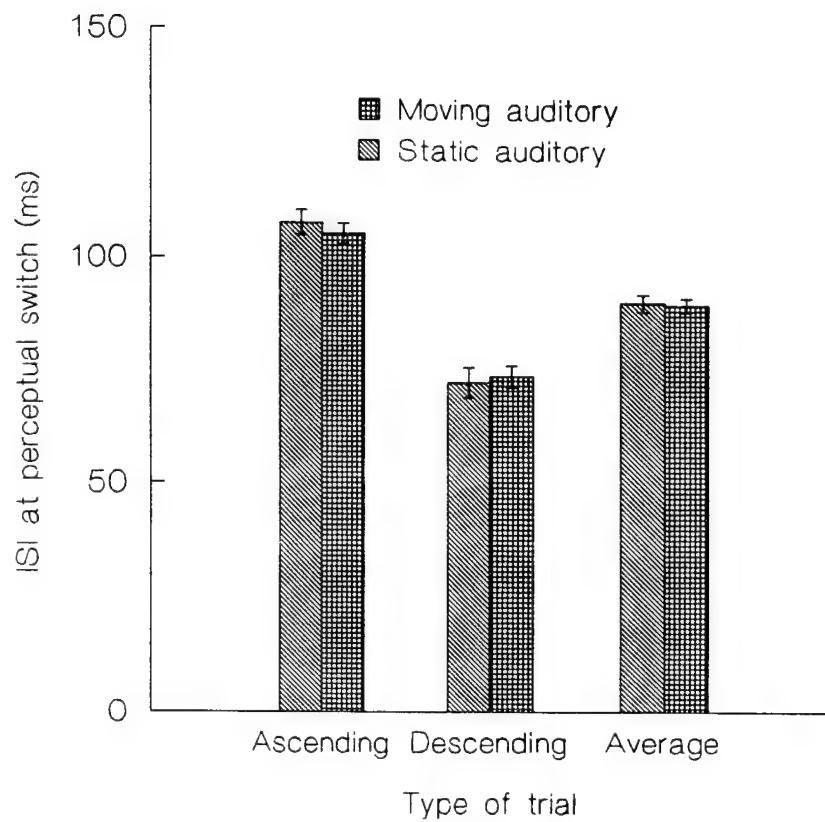


Figure 8.7: The re-categorized threshold ISIs obtained at a horizontal extent of  $3^\circ$  during the ascending trials were significantly affected by the auditory presentation mode.



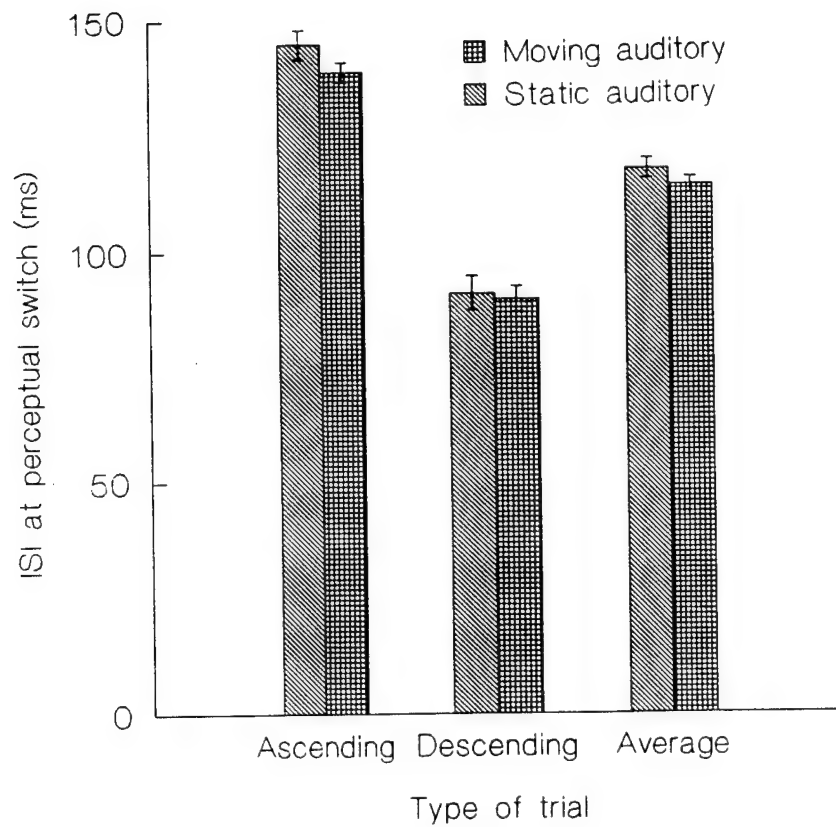


Figure 8.8: The re-categorized threshold ISIs obtained at a horizontal extent of  $9^\circ$  during the ascending trials were significantly affected by the auditory presentation mode.

Source	SS	df	MS	F
$S'_{mode}$	0.71	1	0.71	0.00**
Subject	168880.64	10	16888.06	
Subject x $S'_{mode}$	10253.07	10	1025.31	
$E_H$	37226.94	1	37226.94	16.06*
Subject x $E_H$	23182.09	10	2318.21	
$S'_{mode}$ x $E_H$	203.97	1	203.97	0.58
Subject x $S'_{mode}$ x $E_H$	3487.12	10	348.71	

\*\* Low F-value

\* Significant  $p < .05$

Table 8.5: When the two moving auditory presentation modes were combined into one, only the horizontal extent significantly affected the ISI reported at the perceptual switch from *group* to *element* motion in the descending trials.

Source	SS	df	MS	F
$S'_{mode}$	489.41	1	489.41	0.92
Subject	37271.95	10	3727.19	
Subject x $S'_{mode}$	5313.47	10	531.25	
$E_H$	83194.25	1	83194.25	92.46*
Subject x $E_H$	8997.43	10	899.74	
$S'_{mode}$ x $E_H$	282.31	1	282.31	1.81
Subject x $S'_{mode}$ x $E_H$	1558.49	10	155.85	

\* Significant  $p < .05$

Table 8.6: When the two moving auditory presentation modes were combined into one, only the horizontal extent significantly affected the averaged ISI reported at the perceptual switch from *group* to *element* or *element* to *group* motion.

stimulus extent on the ascending, descending, and  $RL$  threshold  $T_{ISI}$  [ $F(1, 10) = 118.00, p < .05$ ], [ $F(1, 10) = 16.06, p < .05$ ], and [ $F(1, 10) = 92.46, p < .05$ ]. The secondary analysis also reveals a statistically significant effect on the ascending trial threshold  $T_{ISI}$  by the presence of auditory motion [ $F(1, 10) = 6.01, p < .05$ ]. No interaction is indicated between the two independent variables,  $E_H$  and  $S'_{mode}$ .

The reduction of threshold ISI in the moving auditory mode relative to the auditory mode seen in the secondary analysis suggests that the auditory stimulus does aid the long-range motion perception process in attaining the *group* motion perception as the ISI was gradually increased over time and the perception shifted from *element* motion to *group* motion. This effect appears to occur during the ascending trials but not in the descending trials. The absence of interaction between the extent and auditory presentation conditions indicates that this effect occurs within both spatial extents. This effect, seen only within the secondary analysis, is contrary to the experimental hypothesis.

Also of interest is the large separation in the threshold ISIs resulting from the ascending trials relative to the descending trials. The ISI required to elicit a perceptual switch from *element* motion to *group* motion in the ascending trials is much larger than the ISI required to elicit a perceptual switch from *group* motion to *element* motion in the descending trials. This separation is characteristic of a perceptual system incorporating some level of hysteresis. Hysteresis was found to be present by Eggleston [27] in the perception of visual apparent motion.

Taking the results of both the primary and the secondary analyses into perspective, the hypothesis of an auditory stimulus affecting this temporally multi-stable visual display must be accepted. Although the results of the secondary analysis do indicate a  $S'_{mode}$  effect during the ascending trials, no effect is indicated in the descending trials nor on the averaged threshold ISIs. In addition, the initial analysis of variance also does not provide evidence of a  $S_{mode}$  effect.

The secondary analysis does provide very limited support for the trend seen in the averaged data, but taken within the context of the other analyses, does not provide a satisfactory resolution of the conflict between the trends that appear to exist in the data and the initial statistical analysis. By reducing the degrees of freedom of the auditory presentation mode, from two to one, and thus increasing the statistical power of the analysis, the re-categorized analysis yields only one additional statistically significant result. In addition, factors other than statistical power may have affected the experimental data to a greater degree.

An example of these potential factors may be an artifact of the experimental task itself. This artifact may have manifested itself in the analysis of variance results such that the error terms associated with  $S_{mode}$  are much smaller in magnitude than might be expected. This potential artifact may exist in two of the analyses of variance. These two analyses are the secondary analysis of variance summary in Figure 8.5, and the initial analysis of variance summary in Figure 8.2. The relatively small error terms may be a result of how the descending trials were implemented.

For the descending and ascending trials, the computer controller utilized a probabilistic algorithm to adjust the ISI. The probabilistic algorithm was invoked after every set of two frames during the stimulus presentation. Because of this algorithm, when the ISI was small, (the two-frame set sequenced quicker) the probability of the ISI changing within a unit of time was greater than when the ISI was larger (the two-frame set sequence slower). This occurred because the two-frame set took longer to complete for long ISIs than short ISIs.

The experimental results indicate that the shift from *group* motion to *element* motion occurred at relatively short ISIs compared to the shift from *element* to *group* motion. This indicates some overlap in the results from the ascending and descending trials which, in turn, indicates some hysteresis, or organizational capture, in the human motion perceptual system. This hysteresis may mask, in a statistical sense, any threshold shift occurring in the smaller ISIs from the descending trials just due to reaction time influences at the moment of the reported perceptual shift on each trial.

This masking might be reduced by using an experimental task which does not rely on reporting that is prone to reaction time influences. In addition, by using an experimental task which reduces the possibility of hysteresis dominating the results, the potential influence of the moving auditory presentation may be investigated over a broader range of ISI values.

A second experimental investigation was conducted into the potential influence of an auditory stimulus over a temporally multi-stable visual display using an experimental technique conforming to these criteria. This investigation is described in the following section.

## **8.2 Experiment seven: Auditory influence on a visually multi-stable display, a second investigation**

The previous experiment, experiment six, began to characterize an influence of moving auditory stimuli over a temporally driven visual motion perception. However, empirical results from experiment six may have been influenced by hysteresis, or organizational capture, of the human visual system. This experiment, experiment seven, attempts to provide empirical information regarding the characteristics of an auditory influence on temporally multi-stable visual stimuli which would augment results found in experiment six, as well as systematically investigate a broader range of ISI values.

### **8.2.1 Aims**

The purpose of experiment seven was similar to the purpose of experiment six. The purpose was to determine if, and to what extent, the contemporaneous presentation of a moving auditory stimulus may have affected the perceptual organization of a

temporally multi-stable visual display. Specifically, the objective of both experiment six and seven was to evaluate if the temporal threshold, the threshold at which the *element* motion perception and the *group* motion perception would switch, may have been affected by the presence of a linked auditory stimulus. The implications to the model resulting from empirical findings in experiment six were also relevant to empirical findings from experiment seven. Experiment seven differed from experiment six in the experimental methodology utilized. In addition, results from experiment six guided the choice of experimental ranges used for independent variables in experiment seven.

### 8.2.2 Independent variables

There were two independent variables. The first independent variable was the inter-stimulus interval, ISI. The variable ISI had five values in this experiment, 83ms, 100ms, 117ms, 133ms, and 150ms. The second independent variable,  $S_{mode}$ , represented the auditory presentation mode and could represent one of two presentation types, either static or moving.

### 8.2.3 Dependent variable

A single dependent variable was defined for this experiment,  $R_{group}$ . The dependent variable was computed as the percentage of *group* motion reports relative to the total of *group* and *element* motion reports.

### 8.2.4 Other conditions

The auditory signal was presented contemporaneously with the visual three-dot stimulus. The auditory signal became audible with the onset of the visual display but did not become inaudible during the inter-stimulus interval. The three-dot visual pattern was not view-able during the inter-stimulus interval. The audio signal was a noise signal. The spectra of the noise signal had a peak at approximately 80Hz, decreased at approximately 6dB/octave toward lower frequencies, and decreased at approximately 3dB/octave toward the higher frequencies. In addition, at approximately 21kHz, the spectra began to decrease at approximately 25dB/octave.

The auditory signal was presented contemporaneously with the visual stimulus. The auditory signal became audible with the onset of the visual display and did not become inaudible during the ISI. The visual pattern was not view-able during the ISI. Previous work by Strybel *et al* in auditory apparent motion had shown that auditory presentations that become inaudible during the ISI tended to destroy the illusion of motion created by the rapid changing of auditory source position [120].

In the static mode, the auditory signal was presented as a localized source in front of the subject. In the moving mode, the auditory signal was linked with visual

presentation by switching positions coinciding spatially with the left-most dot in the left frame and the right-most during the right frame and remained in that position through the duration and inter-stimulus interval periods. Two auditory presentation modes were utilized from experiment six. The presentation modes utilized were the static auditory mode and the switching auditory mode. Both of these modes are shown in Figure 8.4.

The duration of the visual stimulus during the on periods was 100ms. The ambient sound pressure level within the test booth was 48 dB(A). The sound pressure level within each ear-cup of the headset, when the auditory localizer was positioned at 0° relative to the subject, was 75 dB(A). The attack envelope of the auditory stimulus was 60ms. The decay envelope of the auditory stimulus was 30ms.

The background luminance of the CRT within the booth was  $0.074 \text{ cd/m}^2$  (0.022 Ft-L). The luminance of each dot portrayed on the CRT was  $2.90 \text{ cd/m}^2$  (0.78 Ft-L). The spatial extent of the visual stimulus was  $9.0^\circ$ . The spatial extent was measured from the center of the leftmost dot to the center of the rightmost dot. Each visual dot was square as was  $0.1^\circ$  on each edge. The ISI of the visual stimulus was one of 5 values, either 83ms, 100ms, 117ms, 133ms, or 150ms.

### 8.2.5 Subjects

Eight subjects were used in this experiment. Six subjects were male, 2 subjects were female, and all subjects were between the ages of 18 and 35. Each subject reported no known hearing impairments. Each of the subjects had normal or corrected-to-normal visual acuity. Each subject was a volunteer paid for their time as a subject. Each subject was asked if they had undergone an auditory screening within the last year, and if not, were given an auditory tone test at 250Hz, 500Hz, 1000Hz, 2000Hz, 3000Hz, 4000Hz, and 6000Hz, for both ears. One male subject was found to have a moderate hearing loss in the left ear at 3000Hz and 4000Hz. These particular subject's data were not utilized in subsequent analysis.

### 8.2.6 Task

Each subject was run in one session, with each session consisting of 80 trials each. Combinations of ISI and  $S_{mode}$  were block-randomized forming 8 blocks of 10 trials each. The method of constant stimuli was used.

The trials were self-paced and each lasted approximately 5 seconds. At the beginning of the session, objectives of the experiment and a brief introduction to the experimental set-up were described to the subject. Each subject read and signed a human-use consent form. An auditory tone screening was administered to each subject that indicated they had not had an auditory screening performed within the last year.

A set of training trials was administered following the auditory screening with

feedback given to each subject after each trial. The training consisted of approximately 8 extended duration trials, with an equal number of group motion and element motion presentation made using approximately 17ms ISIs to train the element motion perception and approximately 175ms ISIs to train the group motion perception. The auditory presentation alternated between the static and moving auditory presentation mode during the training trials. The subjects were instructed to report whether they saw group motion or element motion at the end of each presentation. The initial datum collection block was then administered.

Datum collection began with a request of the subject to align the head-tracker. The alignment process determined the subject's head position and attitude within the experimental booth and, thus, supported the centering of the stimulus presentations relative to the subject. Alignment was done by the subject by looking at the center of the monitor and pressing a button inside the subject booth. Once this was accomplished, the monitor displayed a centered fixation target and a surrounding circle filled to indicate the percentage of trials remaining in the block. This remained on the monitor for approximately 3 seconds, and then was replaced with a filled screen for 1 second followed by a blank screen for 1 second. A thin border was then drawn around the edge of the blank screen of the monitor and remained there for 0.5 seconds.

At this time, the head position and attitude was sampled to stabilize the auditory image in the experimental booth. The stimulus was then presented for 3.0 seconds after which the subject was requested to report which motion perception was seen, either *group* motion or *element* motion. The request was made by the computer by removing the visual and auditory display and replacing it with text asking the subject to indicate if they saw *group* or *element* motion. Once the subject made the selection with the joystick, the text was removed and the fixation target was drawn for 1 second and the sequence for the next trial was restarted. A slight break was given to the subjects after 40 trials. The total duration of the session was approximately 1.25 hours.

### 8.2.7 Results and discussion

The number of group motion reports was utilized to calculate a percentage of *group* reports relative to the total number of reports. This percentage was calculated for each combination of ISI,  $S_{mode}$ , and subject. A plot of the percentage of *group* motion reports,  $R_{group}$ , versus ISI averaged across all subjects is shown in Figure 8.9.

An analysis of variance was performed evaluating if  $R_{group}$  was affected by the auditory presentation or the ISI manipulation. The analysis of variance result is summarized in Table 8.7. At an ISI of 83ms, the percentage of *group* reports was 15.77%, and at an ISI of 150ms, the percentage of *group* reports was 67.16%. The ISI significantly affected the percentage of *group* reports, [ $F(4, 24) = 21.07, p < .05$ ]. The auditory presentation also significantly affected the percentage of *group* reports, [ $F(1, 6) = 10.98, p < .05$ ]. This effect can be seen in Figure 8.9. There was no

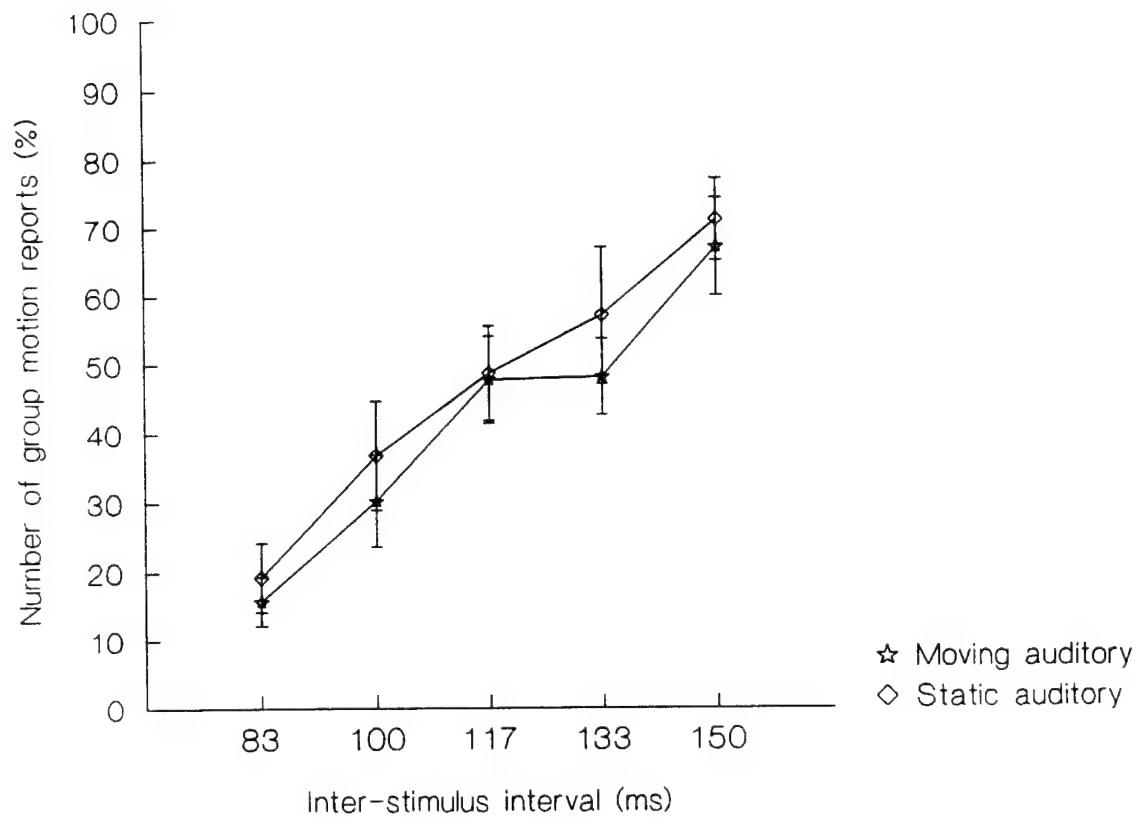


Figure 8.9: The ISI and the auditory presentation mode significantly affected the number of *group* motion reports and no interaction was indicated.



Source	SS	df	MS	F
$S_{mode}$	398.80	1	398.80	10.98*
Subject	10512.28	6	1752.05	
Subject x $S_{mode}$	217.87	6	36.31	
ISI	21513.87	4	5378.47	21.07*
Subject x ISI	6126.17	24	255.26	
$S_{mode}$ x ISI	132.15	4	33.04	
Subject x $S_{mode}$ x ISI	1941.22	24	80.88	0.41

\* Significant  $p < .05$

Table 8.7: The number of group motion reports was significantly affected by the auditory presentation mode and the ISI at which the 3-dot/3-dot display was displayed.

interaction between  $S_{mode}$  and ISI.

These results indicate that both the ISI and the auditory presentation mode affect the percentage of *group* motion reports. No interaction between the effect of ISI and the auditory presentation mode is indicated in the analysis of variance. The lack of interaction between the effect of ISI and auditory presentation mode indicates that the percentage of *group* motion reports can be viewed graphically as two parallel lines, with each line representing one of the auditory presentation modes. These two lines represent small segments of ogive curves which can be visualized to extend and flatten at ISI values greater than 150ms and less than 83ms. The average shift in ISI, indicated as statistically significant from the analysis of variance, is approximately 6.5% from 83ms to 150ms. This shift can clearly be seen in Figure 8.9.

The appearance of a temporal modulation of the influence of the auditory influence over the visual motion perception was indicated in experiment four and suspected in experiment one. A temporal modulation may also be present in results from this experiment. To assess this possibility, the data were analyzed in blocks. Graphs depicting the cumulative percentage of *group* motion reports resulting from each ISI value at each sequential block are shown in Figures 8.10 and 8.11. Figure 8.10 is from the moving auditory presentation trials and Figure 8.11 is from the static auditory presentation trials.

In these graphs, the percentage of *group* motion reports from each value of ISI were stacked on top of one another under each block depicting the relative contribution of each ISI value within each block. The stacking resulted in a maximum level of 500%. There were five values of ISI and thus, if under each ISI value for a particular block every report was a report of *group* motion, the total value under that block would be 500%, coming from the stacking of five 100% bars.

Figures 8.10 and 8.11 appear to show a temporal effect in which a bias toward *group* motion perception was seen in the first block but then disappears in block two through block eight. This temporal effect appears to be consistent in both the

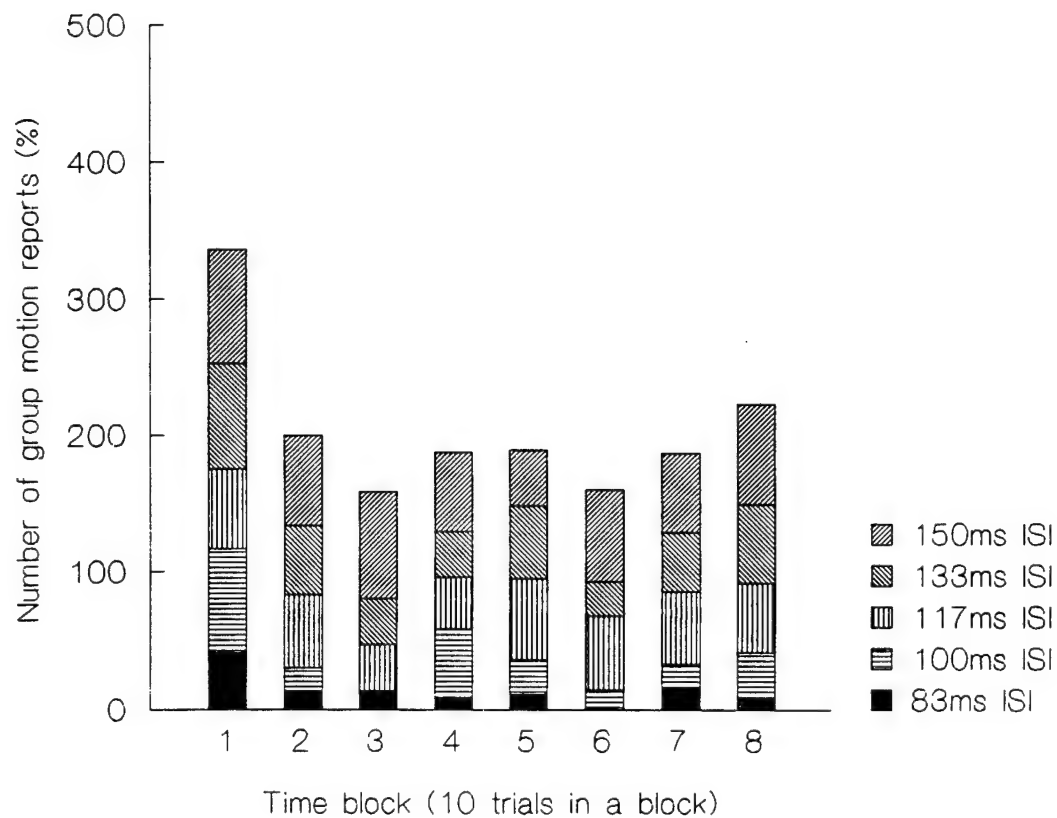


Figure 8.10: The percentage of *group* reports contributed under each of the five ISI levels during the moving auditory presentations was plotted as a function of time. Non-parametric analyses confirmed what appeared to be a temporal affect on the percentage of *group* motion reports as well as indicating that the auditory presentation mode effect was consistent with respect to time.

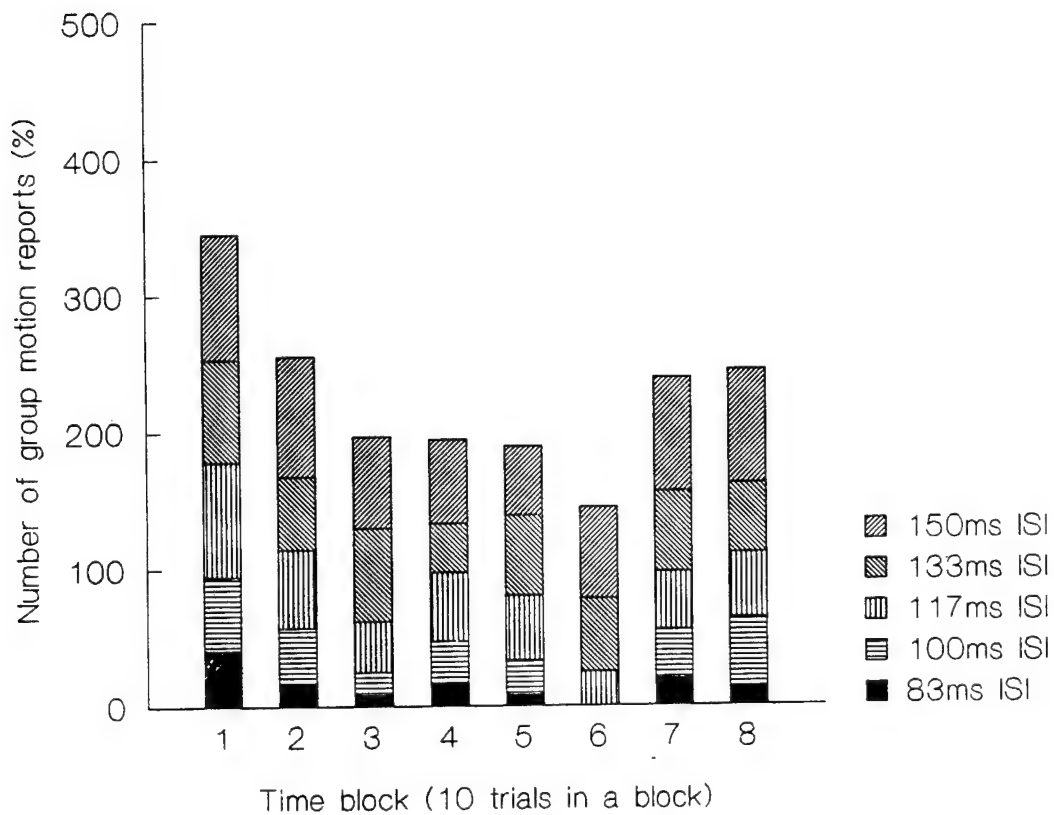


Figure 8.11: The percentage of *group* reports contributed under each of the five ISI levels during the static auditory presentations was plotted as a function of time. Non-parametric analyses confirmed what appeared to be a temporal affect on the percentage of *group* motion reports as well as indicating that the auditory presentation mode effect was consistent with respect to time.

moving and auditory presentations. Figures 8.10 and 8.11 also appear to indicate that the influence of the auditory display on the visual perception is not modulated temporally. The relative increase of  $R_{group}$  in the static auditory trials relative to the moving auditory trials appears not to differ greatly from block to block.

Stating these appearances in a hypothesis form, there are two hypotheses which are to be investigated in a secondary analysis. The first is that the  $S_{mode}$  effect on  $R_{group}$ , as seen in the original analysis of variance, is a function of block. The second is that  $R_{group}$  is a function of block during the trials which incorporate the static auditory presentation and during the trials which incorporate the moving auditory presentation. These two hypotheses form the basis of a secondary analysis.

When the data were separated into cells based on block,  $S_{mode}$ , and ISI, there was only one trial per subject in each cell. Thus, the  $R_{group}$  in each of these cells for each subject was one of two values, either 0.0% or 100.0%. Because of this, statistical verification using an analysis of variance was not an appropriate technique.

Two Friedman two-way analyses of variance were utilized to determine if  $R_{group}$  was a function of the order of presentation. The first analysis used only the trials incorporating the static auditory presentation and the second analysis used only the trials incorporating the moving auditory presentation. For these analyses, the treatment variable was block (the order of presentation), and the case variable was subject-ISI combinations. This technique was appropriate in that the comparisons across blocks were from a single subject at a single ISI and thus, the data across blocks were related. A summary of the results of these two Friedman analyses is shown in Table 8.8 as the leftmost two columns.

Non-parametric statistics were also utilized to determine if the effect of  $S_{mode}$ ,

A third Friedman two-way analysis of variance was utilized to determine if the effect of  $S_{mode}$ , as seen in the original analysis of variance, was a function of presentation order. To determine this, the  $R_{group}$  values from the static trials and moving trials were differenced and the resultant variable was analyzed. The treatment variable in this analysis was block (the order of presentation) and the case variable was subject-ISI combinations. This technique was appropriate in that the comparisons across blocks were from a single subject at a single ISI and thus, the data across blocks were related. A summary of the results of this Friedman analysis is shown in Table 8.8 as the rightmost column.

The Friedman analyses of the trials containing moving and static auditory presentations indicate that the temporal blocking does affect the percentage of group motion reports. This supported the trends seen in Figures 8.10 and 8.11. The Friedman analysis of the *difference* between the static and moving trials indicates that no difference exists in the auditory presentation mode effect throughout the experimental session.

Variable	Static auditory Rank sum	Moving auditory Rank sum	Static and moving auditory Rank sum
Block 1	207.5	206.0	152.0
Block 2	164.5	149.5	170.0
Block 2	151.0	136.5	170.5
Block 4	145.0	148.5	156.5
Block 5	142.0	153.5	149.0
Block 6	123.0	143.0	147.5
Block 7	159.5	152.0	162.5
Block 8	167.5	171.0	152.0
Friedman $F_r$	20.39*	16.08*	2.78

\* Significant at  $p < .05$  assuming a  $\chi^2$  distribution with 7 degrees of freedom

Table 8.8: The Friedman analyses indicated a temporal affect on the percentage of *group* motion reports existed and also indicated that the auditory presentation mode effect was consistent with respect to time.

### 8.3 Contrasts between temporal correspondence influence experimental results

The results from experiment six form an interesting contrast to the results from experiment seven. In experiment six, not enough statistical verification supporting the hypothesized existence of the influence of a moving auditory presentation on the ISIs at which the switch between group and element motion perception occurs was found. One secondary statistical analysis indicated that for ascending ISI trials, the switch from element motion perception to group motion perception occurred at lower ISIs in trials incorporating a moving auditory presentation. In addition, results from experiment six indicated that there is a large difference between the threshold ISIs obtained for the ascending trials relative to the descending trials and indicated some form of hysteresis may have been present.

Based on the results of experiment six, it could be argued that over the range of ISIs between the mean threshold ISIs from the ascending and descending trials, a perceptual bias may have been present which manifests itself as a larger percentage of group motion perception reports occurring with moving auditory presentation than with a static auditory presentation. The support for this belief is weak in that it is based on a bias approximately located between 139ms and 145ms and is supported by only a single, secondary statistical analysis. However, this supposition must be rejected due to the lack of general statistical support.

In contrast to results from experiment six, results from experiment seven clearly

indicate that the moving auditory presentation supports the perception of *element* motion. The effect found in experiment seven is not generally supported in the results from experiment six and is in opposition to the weakly supported bias during ascending ISI trials in experiment six. Clearly there are contrasting results from these two experiments.

Perceptual hysteresis, manifested in experiment six as overlapping means from the ascending and descending trials, is documented in the literature [27]. Perceptual hysteresis may form the basis of an explanation for the contrasting results from experiment six and seven. Experiment six produced threshold ISIs at which perceptual switching occurred. The perceptual switching may occur at the point where perceptual hysteresis breaks down due to characteristics of the visual and auditory stimuli. Obviously, the ISI at which hysteresis may break down may not necessarily be identical to the ISI at which perceptual organization may switch without hysteresis present.

The results from experiment six may indicate that the strength of the perceptual hysteresis is not affected by the presence of the moving auditory presentation relative to the static auditory presentation. The experimental technique utilized in experiment seven reduces the potential for hysteresis by randomizing the ISIs of each stimulus presentation. This is in contrast to experiment six in which the ISI is monotonically adjusted throughout each stimulus presentation. With the reduced hysteresis potential in experiment seven, the ISI at which perceptual organizational switching may occur may be determined more precisely than in experiment six.

## Chapter 9

# Manual tracking of visual-auditory targets

Several temporal and spatial influence characteristics of moving auditory stimuli on visual apparent motion perception were illuminated in the first seven experiments. These seven experiments utilized psychophysical techniques incorporating non-complex visual and auditory stimuli as well as alternative forced-choice reporting by the subjects. This experiment, experiment eight, investigates the possibility that the small *inter*-sensory perceptual influence found in the previous seven experiments may affect the performance of a complex task.

### 9.1 Experiment eight: Manual tracking of an intermittent auditory and visual target

#### 9.1.1 Aims

It was seen in the previous seven experiments that moving auditory stimuli could influence the perceptual organization of visual apparent motion stimuli. This *inter*-sensory influence was small relative to the influence of *intra*-sensory stimulus characteristics and appeared to be significantly reduced by perceptual hysteresis in both spatial and temporal domains.

The inter-sensory influence was evaluated in the previous seven experiments utilizing psychophysical techniques requiring only simple forced-choice responses to be made by the subjects. It is possible that the small *inter*-sensory influence may only be measurable within that type of non-complex environment and may not affect the performance of higher complexity tasks. It can alternatively be proposed that if the auditory influence on visual apparent motion perception is an integral part of the perceptual system, it may influence, to some extent, any task which incorporates visual stimuli having stroboscopic characteristics.

The literature review revealed that human motion perception, in both auditory and visual modalities, appears to be performed at several neurological levels and is both intermediate and central in nature. While some inter-modal effects described within the literature, such as the visual motion interaction with auditory localization, appear to be examples of central-level processing interactions, the majority of literature describing visual-auditory perceptual interactions could be described as processing interactions at the intermediate level within the superior colliculus. As an example of this, performance metrics based on reaction time or attention could be affected by processing within the deep laminae of the superior colliculus and also be multi-modal in nature. This experiment begins to assess the performance implications on manual control of visual-auditory processing at the central and intermediate levels. The manual control task used in this experiment is pursuit tracking of a contemporaneous auditory and visual target.

Two hypotheses form the basis for this experiment. This first experimental hypothesis is that the tracking error measured during pursuit tracking of a pseudo-randomly moving visual target will be reduced by the presence of an auditory target that is linked spatially and temporally with the visual target to be tracked. A graphic representation of a tracking control loop, in which a human is embedded, is shown in Figure 9.1. Within this loop, both visual and auditory stimuli containing information regarding the tracking task are presented to the human. The human processes the information contained in the auditory and visual stimuli and outputs hand movements on a joystick which are feedback to the human as a part of the visual stimuli. This hypothesis is consistent with the premise that the auditory stimuli will enhance target tracking performance by augmenting the visual target position with contemporaneous auditory target position information and thus enhance target motion perception.

The second experimental hypothesis is that the reduction of tracking error resulting from the *inter*-sensory influence will be a function of characteristics of the visual stimulus. While the first hypothesis is somewhat self-defining, the second hypothesis requires further examination.

The second experimental hypothesis anticipates that the *inter*-sensory effect is a function of the characteristics of the visual target movement. This anticipation is based on the results of several of the previous experiments in which the influence of moving auditory stimuli on visual perception was reduced as the perceptual categorization of visual stimuli became less ambiguous. As an example of this, in experiment four, the percentage of vertical motion reports formed an ogive curve when plotted against the horizontal separation of the lower dots of the one-dot/two-dot display. The influence in experiment four could only be measured in the high-slope portion of the ogive curve. This can be seen by comparing Figures 7.9 and 7.10. Another example of this can be observed within the portion of the ogive curve investigated in experiment five which is plotted in Figure 7.11. In Figure 7.11, the mean percentage of vertical motion reports at separations greater than  $5.2^\circ$  was influenced by the moving auditory presentation. In contrast to that result, the mean percentage of vertical motion reports obtained at a separation of  $5.04^\circ$  was not



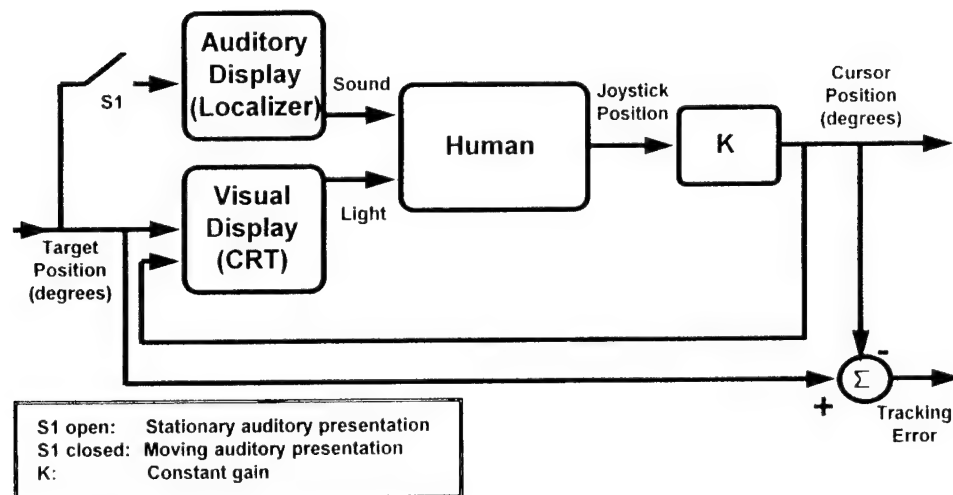


Figure 9.1: The pursuit tracking task displayed to the subject both the cursor and target position. The output of the task could be viewed as either the cursor position alone or the error between the target position and the cursor position.

influenced by the moving auditory presentation.

One potential explanation of these results is that the strength of the *intra*-modal sensation is greater than the strength of the *inter*-modal influence and that only when the *intra*-modal sensation strength is reduced can the effect of the *inter*-modal influence be measured. Based on this premise, the motion perception elicited by a stroboscopic visual target, as well as the manual tracking of that target, may not be measurably influenced by the presence of a linked auditory target if the stroboscopic rate of the visual target is high relative to the bandwidth of the target motion. Alternatively, the motion perception of a stroboscopic visual target, as well as the manual tracking of that target, may be influenced by the presence of a linked auditory target if the stroboscopic rate of the visual target is low.

Within this experiment, narrow-bandwidth pseudo-random visual target movement, sampled at 50 samples per second, was utilized. This sampling rate provided a high stroboscopic rate relative to the bandwidth of the target movement and approximated a continuously moving target. In contrast to the smooth movement, intermittent periods of visual target non-observability were incorporated into the target motion providing gaps of low stroboscopic rate. The combination of the high sample rate of the visual target movement and the intermittent periods provided differing levels of visual target motion predictability.

Several auditory and visual target movement combinations could have been presented during the periods when the *visual* target was non-observable. During these periods, the *auditory* target could have been removed totally, could have been positioned where the visual target would have been if it was observable, could have been moved independently of the non-observable visual target, or could have been made stationary. Previous work by Strybel *et al* in auditory apparent motion had shown that auditory presentations that become inaudible during ISIs much smaller than 400ms tended to destroy the illusion of auditory motion [120]. For this reason, the auditory target could not have been removed totally during the intermittent periods.

The characteristics of the auditory and visual stimulus during the intermittent periods were chosen to reduce additional information regarding target location being presented to the subjects during the intermittent periods and to be consistent with a potential application area of this research. The application area involved the manual tracking of a sensor-derived target which would be sporadically observable. As an example of this application area, the pilot of an aircraft may be required to manually-track a radar return with a cursor that is overlaid on a radar display. The radar cross-section of the object being tracked, as well as radio frequency and environmental interference, may result in intermittent periods of non-observability of the target on the radar display. When the object is non-observable, no information concerning the position of the object to be tracked is available to the radar display.

In this configuration, the target's position and observability can be indicated by the position and presence of the visual target respectively. When the target is observable, the visual target and auditory target are spatially and temporally linked

reflecting the position of the object. When the object is not observable, position information is not updated, the visual target is removed, and the auditory target remains at the last known position of the object. When the object again becomes observable, the visual target and auditory target switch to the new position and are again spatially and temporally linked. In this way, the location of the target is not provided to the subject either visually or aurally during the intermittent periods.

### 9.1.2 Task

Each subject was run in one session that lasted approximately one and one half hour. Each session included a verbal introduction to the overall goals of the research facility, the equipment making up the facility, reading and signing a human-use consent form, training in the performance of the tracking task, and finally collection of data. The collection of data consisted of 4 trials for each subject. The trials were self-paced and each lasted approximately 144 seconds. Each trial incorporated pseudo-random target movement in azimuth interspersed with 400ms periods during which the visual target became non-observable, but continued moving.

Half of the trials were run with stationary auditory stimuli and half were run with auditory stimuli linked spatially and temporally with the target position. However, during the trials using the linked auditory-visual stimuli, when the visual target became intermittently non-observable, the auditory signal became stationary at the last observable target position. When the visual target became observable at the end of the period, the auditory target was switched to the position of the visual target and was again linked spatially and temporally with the visual target.

The intermittent periods were unevenly distributed throughout the 144 second trial yielding some measure of onset unpredictability for the subject. The onset timing of each of the periods relative to the start of the trial is shown in Table 9.1 along with several other potentially salient target movement characteristics.

The task of the subject was to maintain the visual cursor symbol on the visual target symbol as accurately as possible throughout the duration of the trial. Target position and cursor position were recorded throughout each trial. During the intermittent periods, the cursor remained observable and under control of the subject.

The target moved in azimuth only. The time history of the target azimuth movement is shown in Figure 9.2. The target azimuth movement was generated by the sum of 3 sinusoids with frequencies of 0.14Hz, 0.21Hz, and 0.31Hz.

The power spectral density of the target movement was estimated using the Welsh method as implemented by the PC-MATLAB software package [88] using 512 point sections with no overlap. The estimate of the power spectral density is shown in Figure 9.3 with the upper and lower 95% confidence limits shown as the curves above and below the spectral estimate. The noise floor of this spectral estimate appeared to be located at approximately  $10^{-5} \frac{\text{deg}^2}{\text{Hz}}$ .

Period Number	Time at onset (seconds)	Position at onset (°)	Velocity at onset (°/second)	Acceleration at onset (°/second <sup>2</sup> )	Position at end (°)	Velocity at end (°/second)
1	3.42	6.35	6.28	-14.50	7.32	-0.57
2	8.14	10.86	-5.71	-20.17	7.08	-10.57
3	15.26	-3.60	-5.15	19.01	-3.90	2.57
4	19.0	-12.50	-8.29	29.52	-13.38	2.29
5	24.60	-3.60	-11.42	1.72	-7.90	-8.85
6	30.50	10.32	5.63	-26.76	10.50	-3.72
7	37.08	10.80	-4.85	-25.79	6.72	-12.85
8	43.08	0.0	11.43	3.11	4.62	10.28
9	48.92	-4.62	12.57	19.70	1.92	17.15
10	54.82	-1.32	-6.85	2.76	-3.72	-4.28
11	61.32	-10.68	-11.72	18.05	-13.74	-3.72
12	64.16	12.78	-6.00	-26.82	8.34	-12.85
13	70.90	1.14	-6.85	-1.64	-1.50	-5.43
14	74.98	-12.72	-11.15	28.83	-14.64	-0.28
15	82.52	-1.26	11.43	-0.16	3.18	9.15
16	87.72	1.62	-16.85	-4.72	-5.28	-15.43
17	94.04	-1.74	-14.85	0.30	-7.44	-12.28
18	100.20	9.78	2.29	-25.59	8.64	-5.72
19	127.24	5.16	-16.57	-10.98	-2.40	-17.72
20	113.48	6.24	1.72	-16.75	5.70	-2.85
21	117.84	-12.00	-2.00	27.91	-10.20	8.29
22	126.18	3.18	4.57	-11.99	3.96	0.00
23	132.68	3.48	18.28	-13.99	9.60	11.43
24	137.46	-5.76	-4.28	10.17	-6.42	0.28

Table 9.1: Twenty-four 400ms periods, when the visual target was non-observable, were interspersed throughout each trial. The characteristics of the target motion, such as position and velocity, before and after each period, were different from one another.

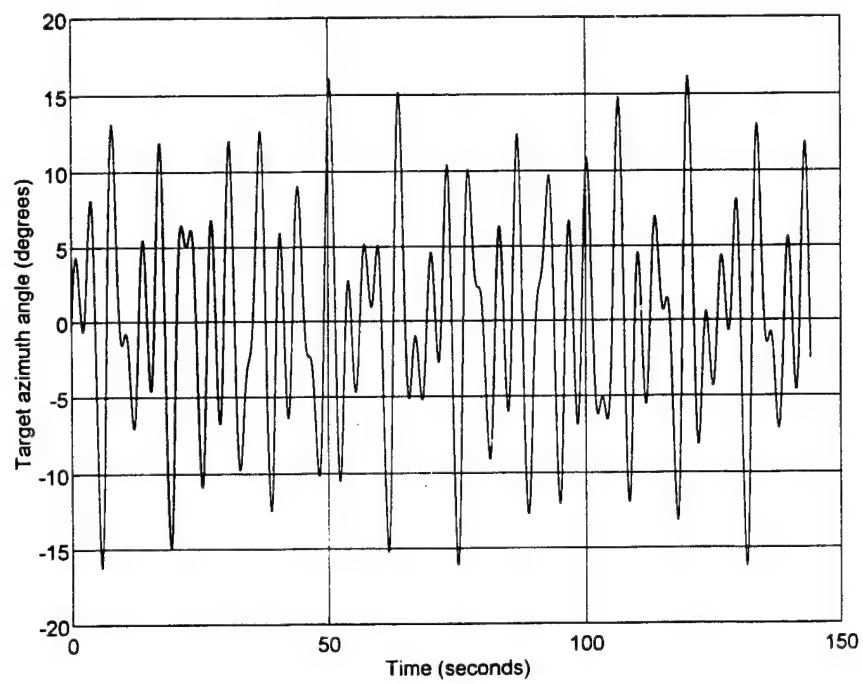


Figure 9.2: The azimuth angle of the visual target changed in a pseudo-random fashion during the 144 second trial.

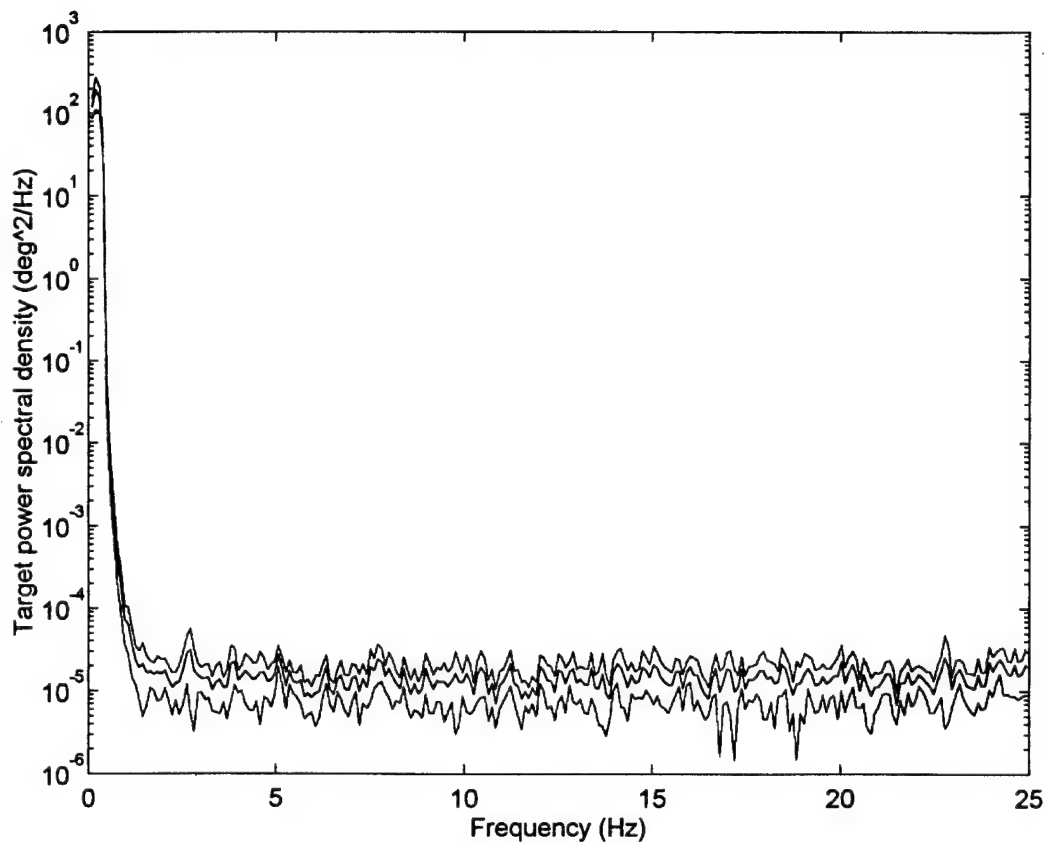


Figure 9.3: The pseudo-random pattern of target azimuth movement was generated as the sum of 3 sinusoids with frequencies of 0.14Hz, 0.21Hz, and 0.31Hz. The upper and lower 95% confidence limit curves are shown above and below the spectral estimate curve. The resolution of this power spectral density estimate is approximately 0.1Hz.

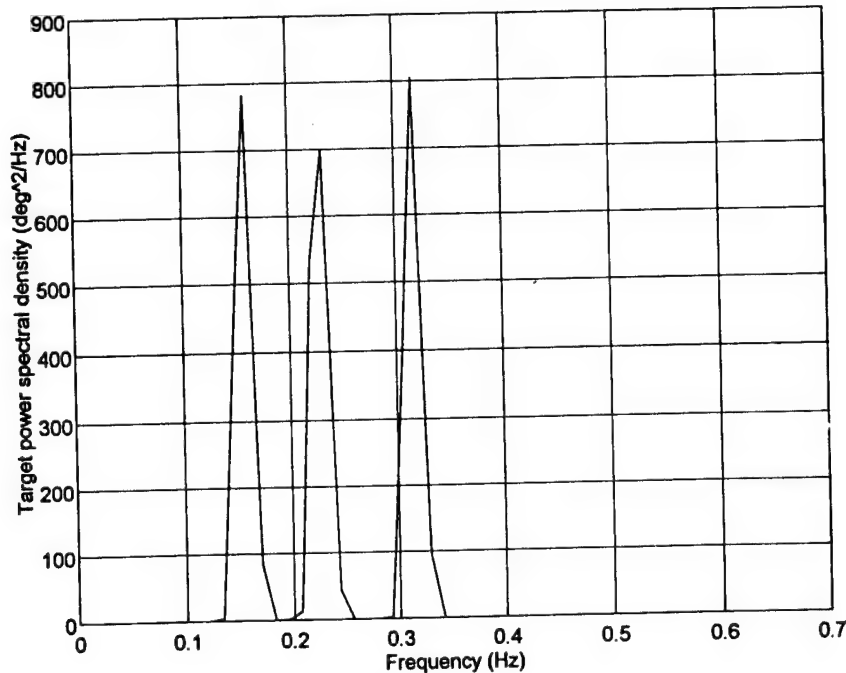


Figure 9.4: The three sinusoids generating the target motion can be recognized in a power spectral density estimate using 4096 points which yields a resolution of approximately 0.01Hz.

A higher resolution estimate at the lower frequencies was obtained utilizing the Welch method with a 4096 point section. This estimate is shown in Figure 9.4.

Another description of the target movement can be postulated. During the intermittent periods, the position of the visual target can be considered undefined because it is not observable. The position of the auditory source, during the intermittent periods, remained at the last observable position of the visual target. The second description of the target movement can be formulated by maintaining the target position at the last observable position of the visual target during each of the intermittent periods. In essence, the time history of the target position can be considered to be *latched* to the last observable position of the visual target during each of the intermittent periods. This latching increases power in the target movement power spectral density at frequencies higher than the three sinusoids. The power spectral density of this second target movement formulation is shown in Figure 9.5.

Training consisted of instructing the subject on how to use the stick to control the cursor and allowing them to practice the task, both with moving and stationary auditory stimuli. A score based on the subject's error for the trial was displayed after each practice trial. Training was continued until the subject performed three sequential trials with r.m.s. tracking errors within 7% of one another. The first set in

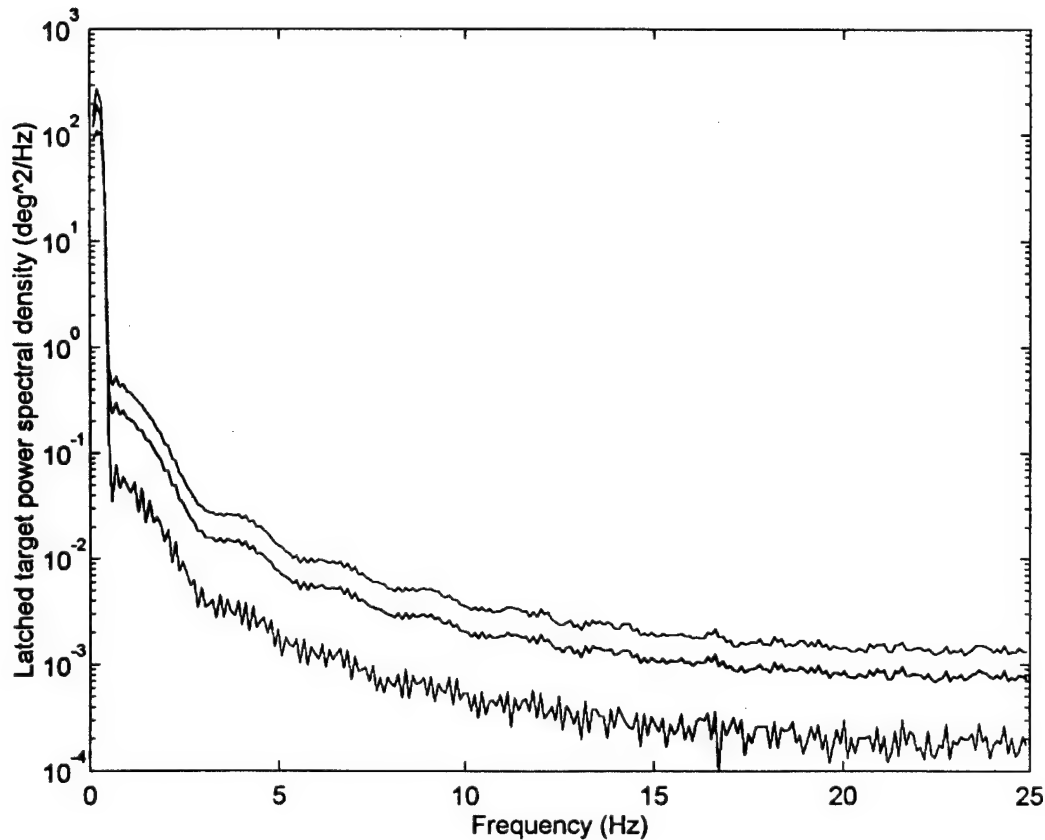


Figure 9.5: The pseudo-random pattern of target azimuth movement was generated as the sum of 3 sinusoids with frequencies of 0.14Hz, 0.21Hz, and 0.31Hz and latching the target position to the last observable target position of the visual target during the intermittent periods. The upper and lower 95% confidence limit curves are shown above and below the spectral estimate curve. The resolution of this power spectral density estimate is approximately 0.1Hz.



which data were collected was administered after the training was completed.

### 9.1.3 Independent variables

There was a single independent variable utilized within this experiment. This was the auditory presentation mode,  $S_{mode}$ .  $S_{mode}$  could have taken one of two values within each trial, either stationary (S) or moving (M).

In the stationary mode ( $S_{mode} = S$ ), the auditory signal was presented as a localized, stationary source in front of the subject. In the moving mode ( $S_{mode} = M$ ), the auditory signal was linked spatially and temporally outside of the non-observable periods of the visual target.

During the periods when the visual target was non-observable under the  $S_{mode} = M$  condition, the auditory target remained stationary at the last observable target location until the visual target again became observable, at which time the auditory target switched to the location of the visual target and was again linked spatially and temporally with the visual target. In this way, the location of the target was not provided to the subject either visually or aurally during the intermittent periods.

### 9.1.4 Dependent variable

A single type of dependent variable was defined for this experiment. The dependent variable was the error between the cursor and the visual target to be tracked. The cursor position was controlled by the subject and the target position was controlled by the computer. During the intermittent periods, when the visual target was non-observable and the auditory target was stationary, the position of the visual target used to calculate error was the position of the visual target if it had been observable. The error was calculated as root-mean-square (r.m.s.) degrees.

### 9.1.5 Other conditions

The visual symbology was portrayed with a luminance of  $2.5 \text{ cd/m}^2$ , and shown against a dark background of  $0.14 \text{ cd/m}^2$ . The target diamond subtended a visual angle of  $1.2^\circ$  horizontally and vertically. The visual cursor, an open-center cross-hair, subtended an angle of  $2.1^\circ$ . In a zero error condition, the points of the diamond were adjacent to the inside edges of the cursor. The visual target tracking symbology is shown in Figure 9.6.

The ambient noise level in the experimental area was 46 dB(A). The audio signal was a noise signal generated by a B&K audio noise generator, type 1405. The auditory level of the noise was 65 dB(A). Sound levels were measured with a B&K Precision Sound Level Meter, type 2235, with a type 4176 microphone insert. Sound levels measured inside the headphones were measured with the same meter combined with an artificial ear and 4144 microphone insert at a  $0^\circ$  localizer azimuth position.

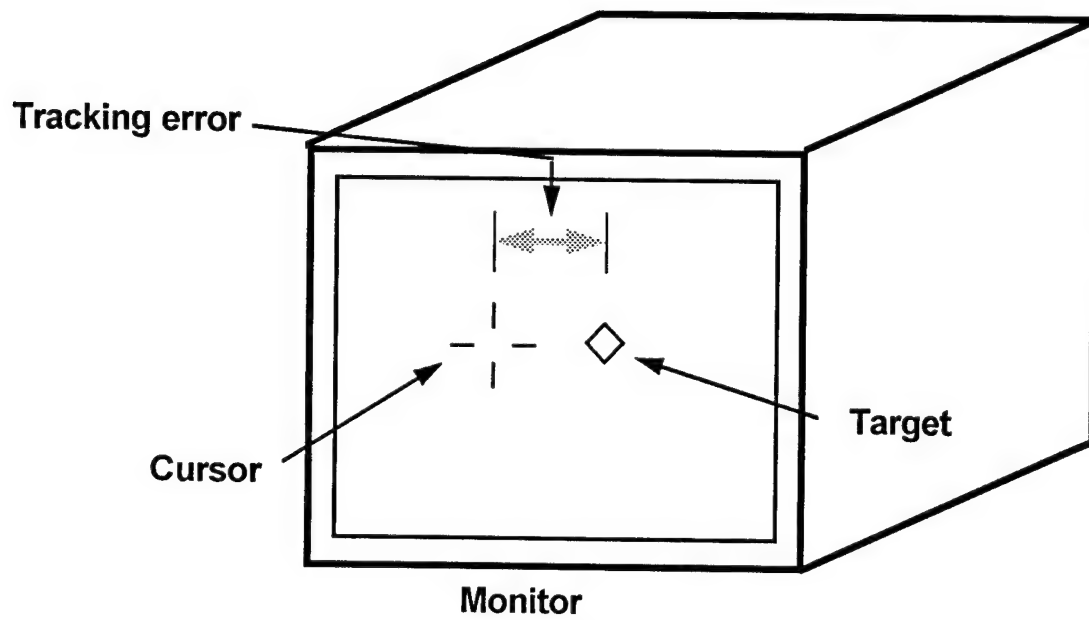


Figure 9.6: The subject's task was to keep the cursor centered on the target throughout the trial using a joystick. The joystick position was transformed into the cursor azimuth angle by multiplication of a constant.

The spectra of the noise signal had a peak at approximately 80Hz, decreased at approximately 6dB/octave toward lower frequencies, and decreased at approximately 3dB/octave toward the higher frequencies. In addition, at approximately 21kHz, the spectra began to decrease at approximately 25dB/octave.

### 9.1.6 Subjects

Seven subjects were used in this experiment. Five were male, two were female, and all were between the ages of 18 and 38. Each subject reported no known hearing impairments. Each of the subjects had normal or corrected-to-normal visual acuity. Each subject was a volunteer paid for their time as a subject.

### 9.1.7 Results and discussion

This section describes the analyses and results obtained from data generated within experiment eight.

The r.m.s. tracking error data were organized by taking 5 sequential samples of data, each sample being 20 seconds in duration, from each subject in each trial. The selection of the number of samples was arbitrary. These samples were obtained from the central portion of each run such that tracking error data at the start of a run and at the completion of each run were not considered. From the seven subjects, this generated 140 samples of performance, 70 from the static auditory presentation condition and 70 from the moving auditory presentation condition. A plot of the mean r.m.s. error between the target and the cursor, for the two auditory conditions, is shown in Figure 9.7.

The trend in the data of Figure 9.7 suggests that the mean tracking error is slightly reduced in the presence of moving auditory stimuli relative to the tracking error in the presence of static auditory stimuli implying that performance was enhanced under the moving auditory presentation relative to the static auditory presentation.

To investigate this apparent trend, an analysis of variance was performed on the samples of tracking error. A summary of the analysis of variance result is shown in Table 9.2. The analysis of variance does not support the trend of increased performance in the moving auditory stimulus condition as it indicates that there is no significant effect of the auditory presentation mode on the tracking error.

In addition to overall tracking performance, tracking performance associated with the intermittent periods is also of interest. However, the performance associated with the intermittent periods may be masked by long-duration measures of tracking performance, such as a 20-second r.m.s metric, because the intermittent periods occupy only a fraction of the total trial duration. To analyze the tracking performance associated with the intermittent periods, tracking error was computed for each trial, subject, and intermittent period under each auditory presentation mode both during the intermittent period and the 400ms following the intermittent

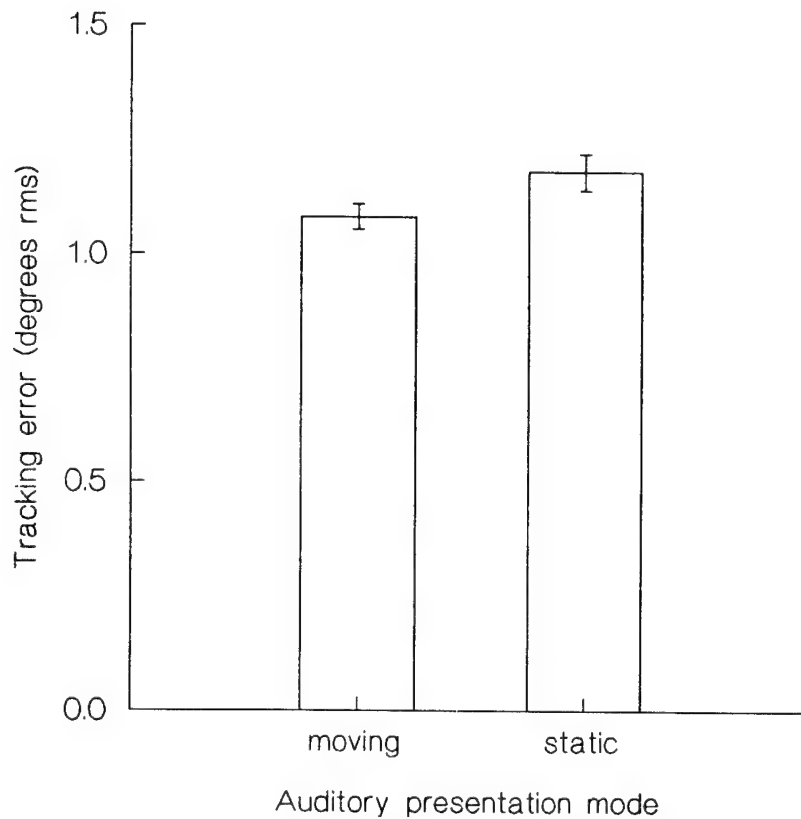


Figure 9.7: The mean r.m.s. tracking error, obtained as the average of the r.m.s. error from each trial, was not significantly affected by the auditory presentation mode.

Source	SS	df	MS	F
$S_{mode}$	0.34	1	0.34	2.53
Subject	6.97	6	1.16	
Subject x $S_{mode}$	0.81	6	0.14	
Sample	0.02	4	0.005	0.12
Subject x Sample	0.99	24	0.041	
$S_{mode}$ x Sample	0.125	4	0.031	0.816
$S_{mode}$ x Sample x Subject	0.912	24	0.038	

Table 9.2: An analysis of variance indicates that the auditory presentation mode, represented by  $S_{mode}$ , does not significantly affect the tracking error. The tracking error time histories were separated into sequential 20-second time-slices. Sample, in this table, is the number of the time-slice. The tracking error time-slice does not significantly affect the tracking error and no interaction between the auditory presentation and the time-slice is indicated.

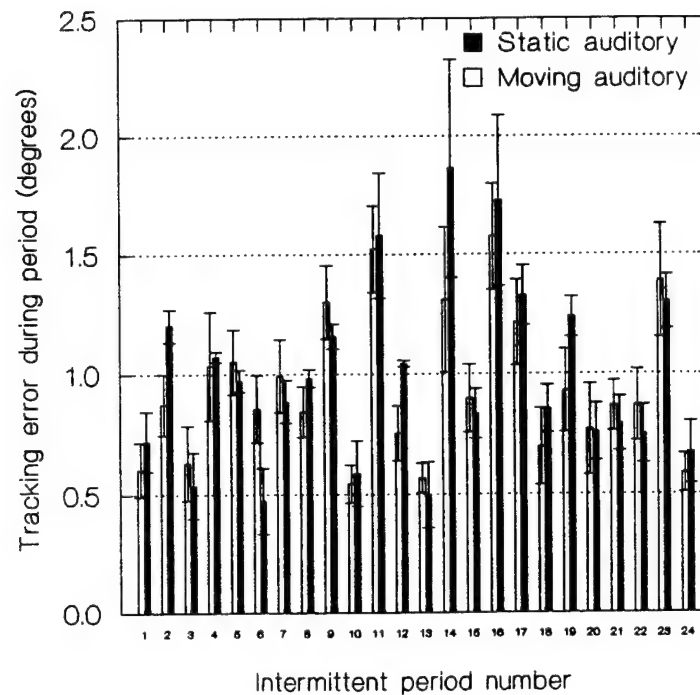


Figure 9.8: The tracking errors obtained during the intermittent periods from all subjects and trials were plotted.

period. A graphic depiction of the mean of the r.m.s. tracking errors were plotted against period number for both the moving and stationary auditory trials. These plots are depicted in Figures 9.8 and 9.9.

It appears, from observing Figures 9.8 and 9.9, that the variance associated with each period varies greatly across periods. This may have been due to target movement characteristics combined with characteristics of the human's manual control capabilities. Two analyses of variance, examining the effect of the auditory presentation mode and period number, were performed on the tracking error during the intermittent periods and the 400ms following the end of the intermittent periods. The results of these analyses of variance are depicted in Tables 9.3 and 9.4.

Target movement characteristics associated with the individual intermittent periods affects tracking performance. This is substantiated by the analyses of variance which shows a significant effect of period on tracking error during the intermittent periods [ $F(23, 138) = 5.31, p < .05$ ] and during the 400ms interval following the intermittent periods [ $F(23, 138) = 6.52, p < .05$ ]. However, the periods differ in several characteristics and because of this, it is not possible to determine from these analyses which particular characteristic may be the cause of the performance effect.

No clear *inter*-sensory effect on manual tracking appears to exist in these data based upon the statistical analyses. Indeed, it may be that the small perceptual effects that were illuminated in the previous seven experiments do not transfer into performance effects in this manual control task. Conversely, performance effects may

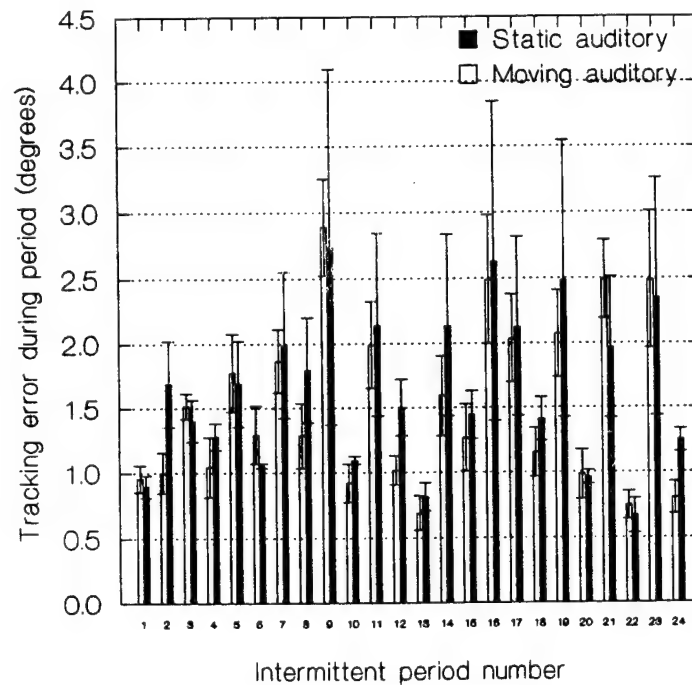


Figure 9.9: The tracking errors obtained during the 400ms following the intermittent periods from all subjects and trials were plotted.

Source	SS	df	MS	F
$S_{mode}$	0.38	1	0.38	0.76
Subject	31.13	6	5.19	
Subject x $S_{mode}$	3.02	6	0.50	
Period	66.61	23	2.98	5.31*
Subject x Period	77.40	138	0.56	
$S_{mode}$ x Period	6.11	23	0.27	0.56
Subject x $S_{mode}$ x Period	66.06	138	0.48	

\* Significant  $p < .05$

Table 9.3: An analysis of variance was performed on r.m.s. tracking performance during the intermittent periods. In this analysis, the auditory presentation mode did not affect the tracking error during the intermittent period. However, some characteristic, or characteristics, of the periods themselves did significantly affect tracking performance.

Source	SS	df	MS	F
$S_{mode}$	2.85	1	2.85	0.98
Subject	54.15	6	9.03	
Subject x $S_{mode}$	17.43	6	2.91	
Period	232.64	23	10.11	6.52*
Subject x Period	213.93	138	1.55	
$S_{mode}$ x Period	13.18	23	0.57	0.54
Subject x $S_{mode}$ x Period	146.47	138	1.06	

\* Significant  $p < .05$

Table 9.4: An analysis of variance was performed on r.m.s. tracking performance during the 400ms interval following the intermittent periods. In this analysis, period significantly affected tracking performance during the 400ms interval following the intermittent periods. The auditory presentation mode did not significantly affect tracking performance during the 400ms interval following the intermittent periods.

be present but not observable within the tracking error data due to their magnitude relative to other factors affecting the tracking error data. Analyzing the tracking error from a frequency and systems perspective, rather than a time domain perspective, may enhance the clarity of the experimental results.

### Frequency domain analyses

The analyses of variance of r.m.s. tracking error depicted in Tables 9.3 and 9.4 are analyses of time domain performance characteristics. An alternative type of analysis was performed, specifically a frequency domain type of analysis, in an attempt to further refine the characteristics of the target movement that appear to affect the difference in tracking performance between the moving and static auditory presentations.

In the frequency domain analysis, the subject was viewed as a part of a system, with the input being the target position, as a function of time, and the output being the cursor position, as a function of time. The system also incorporated feedback to the subject of the cursor position. The feedback came from the visual display which depicted both the target position as well as the cursor position. A block diagram depicting this pursuit tracking task is shown as Figure 9.1.

To begin this analysis, the power spectral density of the cursor position, representing an output of the system, was estimated for both the static auditory presentation trials and the moving auditory presentation trials. The output power spectral density estimates were computed by estimating the power spectral density of each trial individually, separating the static auditory presentation trials from the moving auditory presentation trials, and averaging the resultant ensembles into two power spectral densities, one representing the static auditory presentation trials and

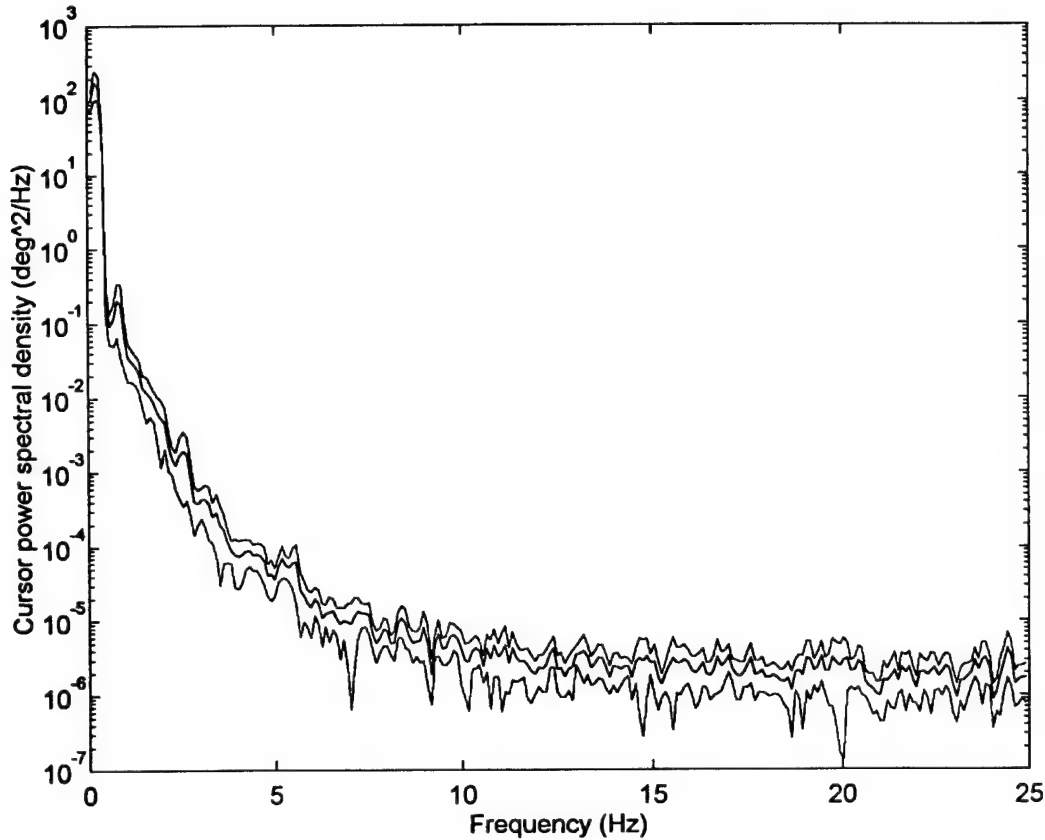


Figure 9.10: The cursor power spectral density estimate, from the static auditory trials, contains significant power at frequencies lower than approximately 2Hz. The upper and lower 95% confidence limit curves are shown above and below the spectral estimate curve. The resolution of this power spectral density estimate is approximately 0.1Hz.

a second representing the moving auditory presentation trials.

The power spectral density of the cursor movements was estimated using the Welch method as implemented by the PC-MATLAB software package [88] using 512 point sections with no overlap. These two power spectral density plots are shown in Figures 9.10 and 9.11 with the upper and lower 95% confidence limits shown as the curves above and below the spectral estimate. The noise floor of these spectral estimates appeared to be located at approximately  $10^{-6} \frac{\text{deg}^2}{\text{Hz}}$ .

Figures 9.10 and 9.11 appear to be very similar to one another. Within Figures 9.10 and 9.11, there appears to be more power in the frequencies below 2Hz, than is present in the target movement, as depicted in Figure 9.3. However, Figures 9.10 and 9.11 appear to correspond more closely with Figure 9.5, the power spectral density of the target movement *with* latched positions during the intermittent periods, than with Figure 9.3, the power spectral density of the target



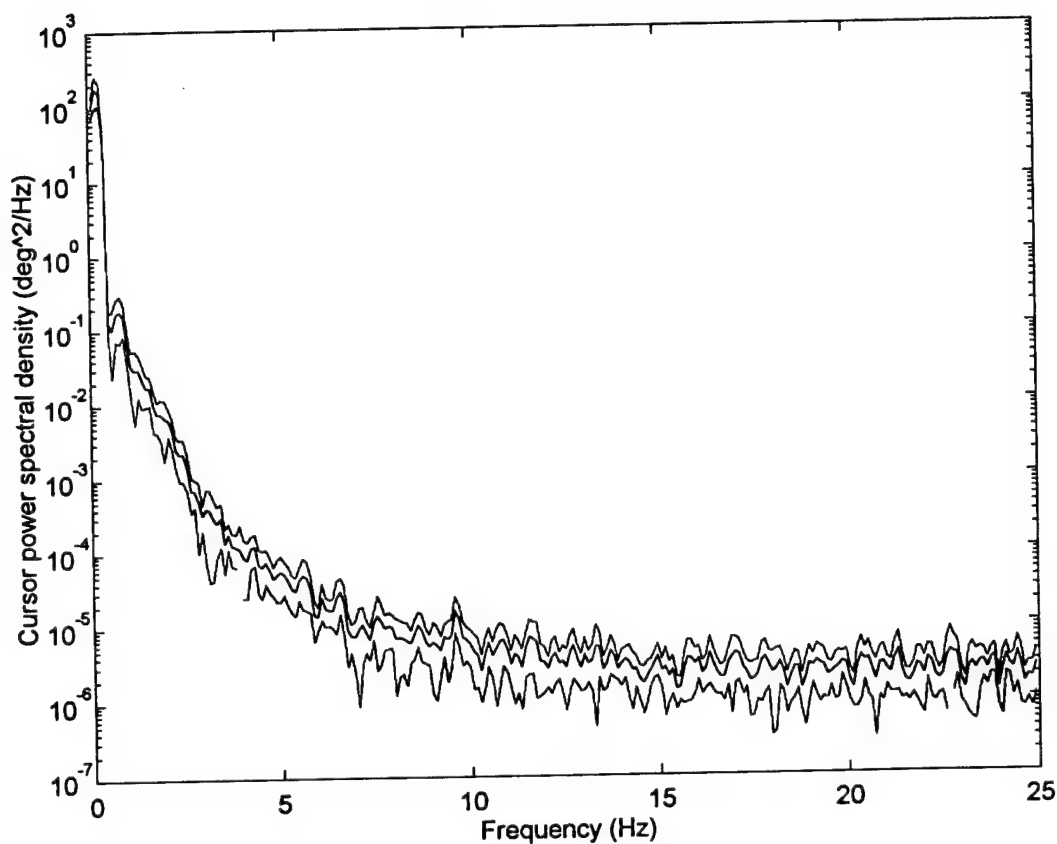


Figure 9.11: The cursor power spectral density estimate, from the moving auditory trials, contains significant power at frequencies lower than 2Hz. The upper and lower 95% confidence limit curves are shown above and below the spectral estimate curve. The resolution of this power spectral density estimate is approximately 0.1Hz.

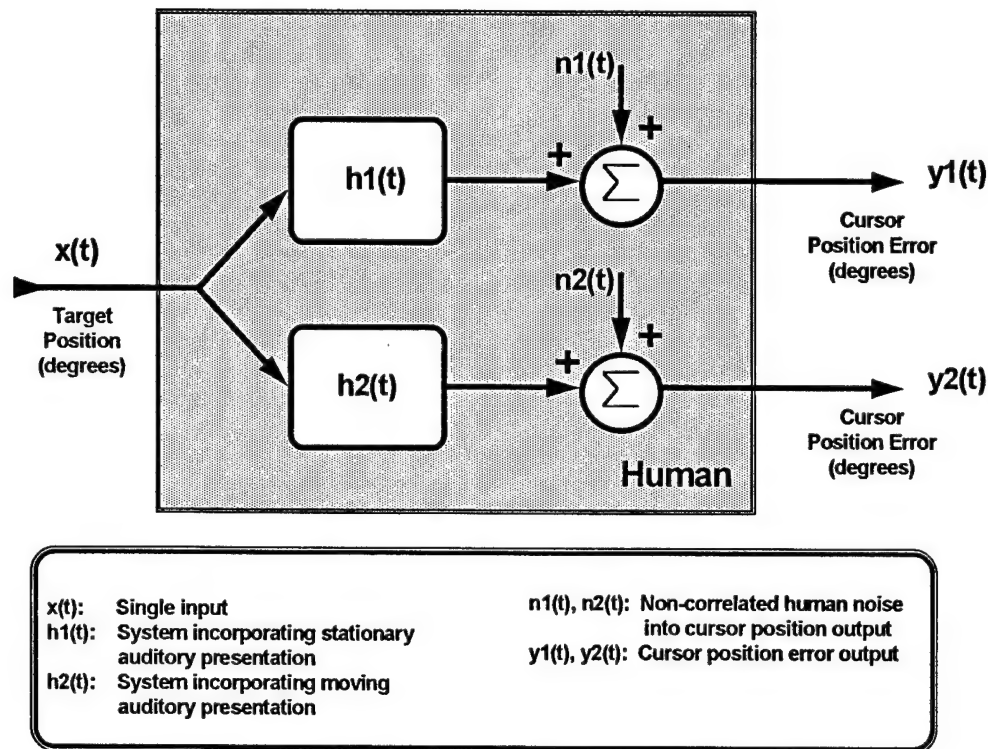


Figure 9.12: A parsimonious model that supported methodical systems analysis of the tracking task was a single input/multiple output model with additive noise at the output.

movement *without* latched positions during the intermittent periods.

To begin to investigate these observations, the block diagram shown as Figure 9.1 was simplified into a parsimonious single input/multiple output representation that supported the estimated spectral relationship between the target position as input and the cursor position error as output. This model is shown as Figure 9.12. In this representation, the moving and non-moving auditory presentation modes were modeled as two different systems, and the output of the system was composed of linear responses to the input signal and additive noise components. The additive noise component was necessary to model the significant power at frequencies above the bandwidth of the target movement that appear to exist in the cursor power spectral density. The linear systems, in combination with the additive noise components, formed two models of the human.

The power spectral densities of the correlated and non-correlated errors for each trial were estimated for both the moving and static presentation models using the target movement described in Figure 9.3. The estimation of the power spectral

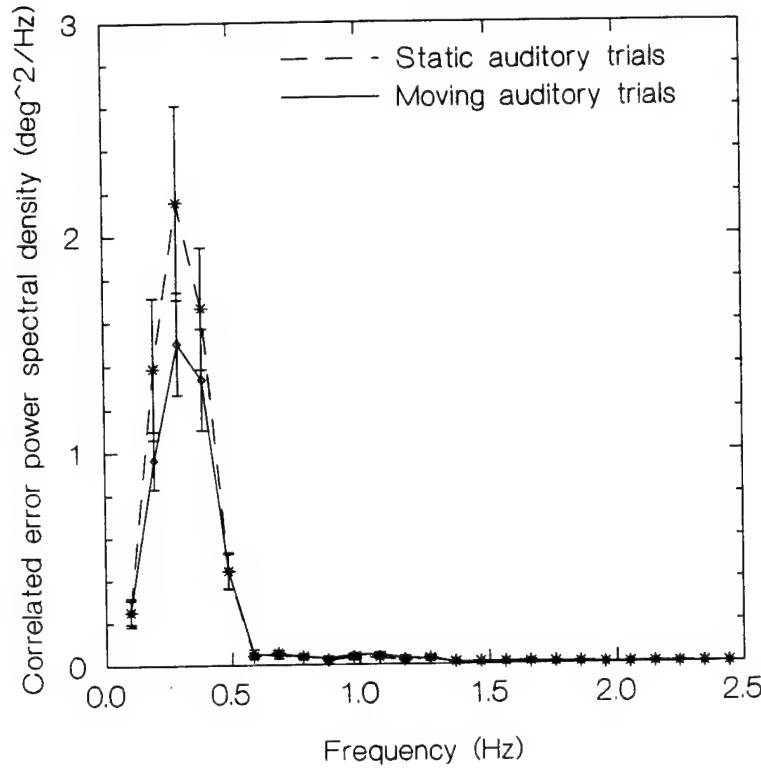


Figure 9.13: The power spectral density of the correlated error within each moving and static auditory trial from each subject was estimated. The resultant estimates were averaged across frequency components. Standard error is indicated at each frequency component. The estimates were based on target movement shown in Figure 9.2. The resolution of this power spectral density estimate is approximately 0.1Hz.

density was performed using the models found in Figure 9.12 and the equations shown below taken from Bendat [10]. In these equations,  $G_{xx}(f)$  is the one-sided autospectral density function of the input  $x(t)$ ,  $G_{yy}(f)$  is the one-sided autospectral density function of the output  $y(t)$ , and  $G_{xy}(f)$  is the one-sided cross-spectral density function between  $x(t)$  and  $y(t)$ . The correlated error power spectral density is represented by  $G_{vv}(f)$  and the non-correlated error power spectral density is represented by  $G_{nn}(f)$ .

$$G_{vv}(f) = \left| \frac{G_{xy}(f)}{G_{xx}(f)} \right|^2 G_{xx}(f) \quad (9.1)$$

$$G_{nn}(f) = G_{yy}(f) - G_{vv}(f) \quad (9.2)$$

The estimated power spectral densities of the correlated and non-correlated errors from each trial were averaged at each frequency component. The estimated power spectral densities of the correlated and non-correlated errors, along with associated standard error bars, are shown in Figures 9.13 and 9.14.

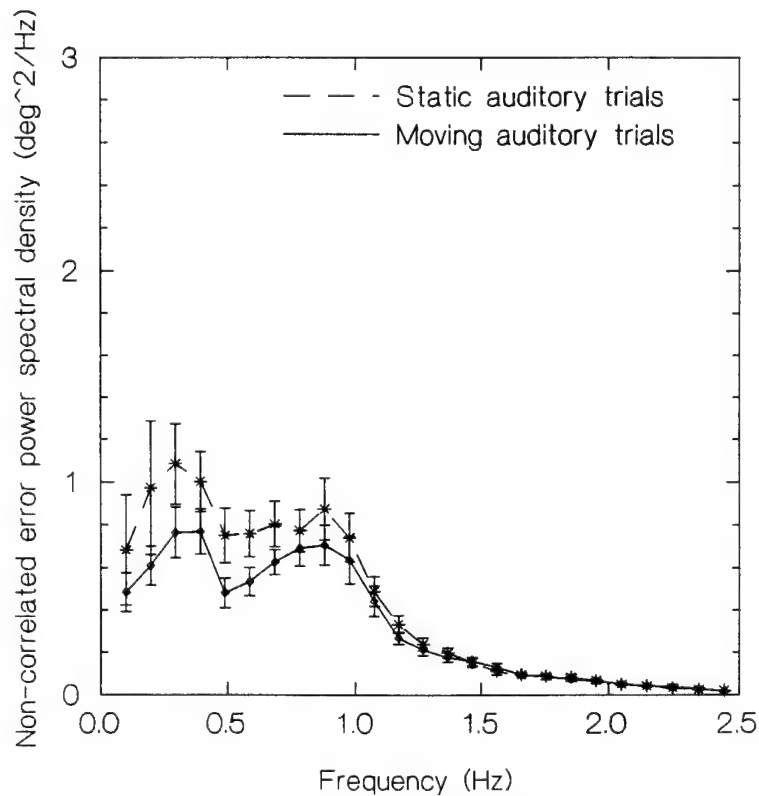


Figure 9.14: The power spectral density of the non-correlated error within each moving and static auditory trial from each subject was estimated. The resultant estimates were averaged across frequency components. Standard error is indicated at each frequency component. The estimates were based on target movement shown in Figure 9.2. The resolution of this power spectral density estimate is approximately 0.1Hz.

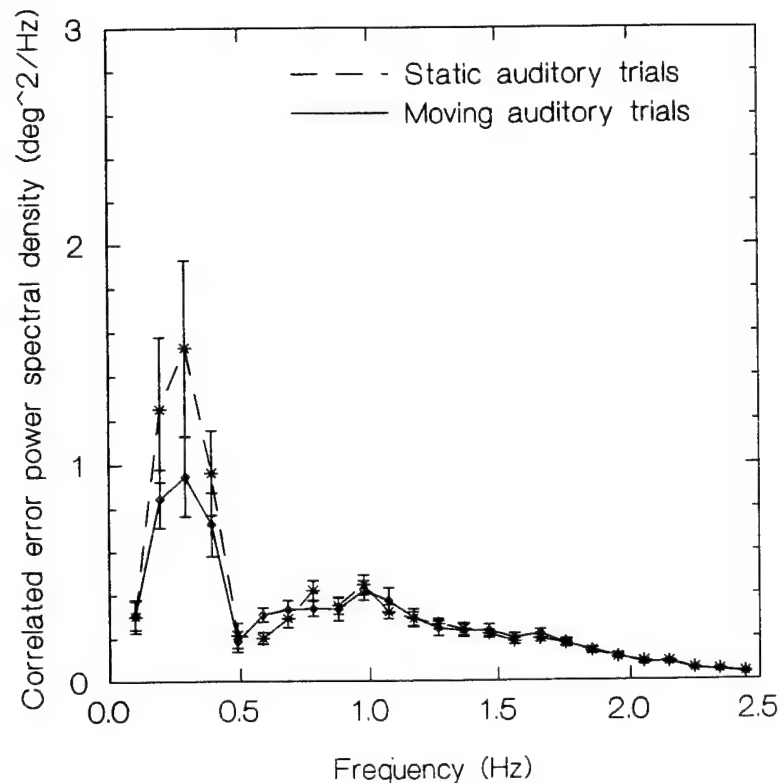


Figure 9.15: The power spectral density of the correlated error within each moving and static auditory trial from each subject was estimated. The resultant estimates were averaged across frequency components. Standard error is indicated at each frequency component. Target movement included latched intermittent periods. The resolution of this power spectral density estimate is approximately 0.1Hz.

The power spectral densities of the correlated and non-correlated errors were also estimated for both the moving and static presentation models using the target movement described in Figure 9.5. The estimation of the power spectral density was again performed using the models found in Figure 9.12 and the algorithm described by Bendat [10] and shown above. The estimated power spectral densities of the correlated and non-correlated errors from each trial were averaged at each frequency component. The estimated power spectral densities of the correlated and non-correlated errors, along with standard error bars, are shown in Figures 9.15 and 9.16.

The figures 9.16, 9.15, 9.14, and 9.13 imply that a reduction in correlated error under the moving auditory presentation mode, relative to the static auditory presentation mode, exists in the frequency range from 0.1Hz to 0.5Hz and that a reduction in non-correlated error under the moving auditory presentation mode, relative to the static auditory presentation mode, exists in the frequency range from 0.1Hz to 1.0Hz. The data at each frequency component are organized as paired comparisons between the moving auditory presentation mode and the static auditory presentation mode for each subject. An analysis using the Wilcoxon signed-rank test was utilized to determine if a statistically significant difference existed between the

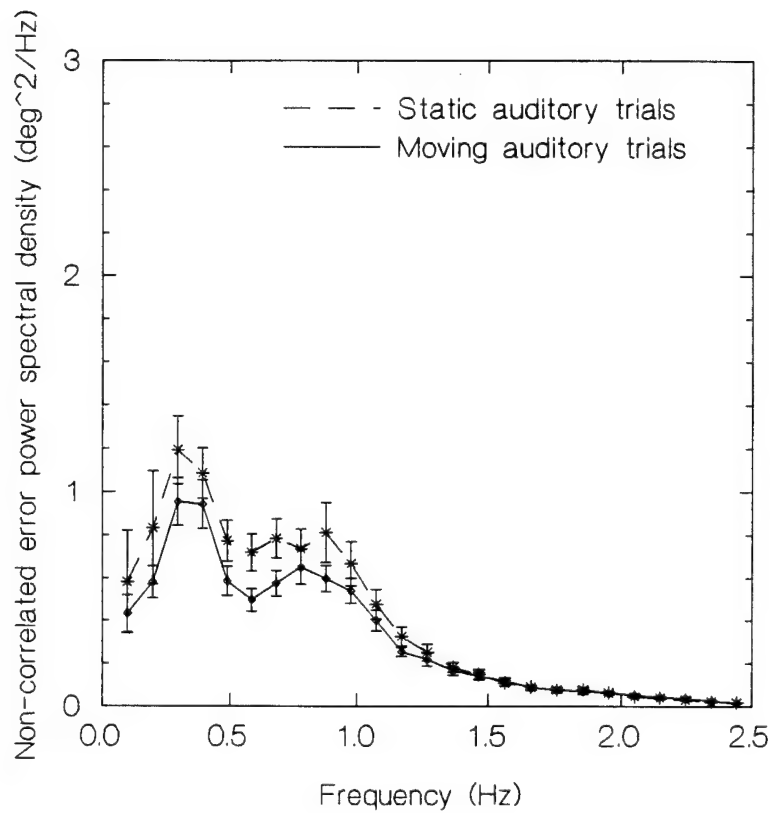


Figure 9.16: The power spectral density of the non-correlated error within each moving and static auditory trial from each subject was estimated. The resultant estimates were averaged across frequency components. Standard error is indicated at each frequency component. Target movement included latched intermittent periods. The resolution of this power spectral density estimate is approximately 0.1Hz.

Frequency range	Target description	Error estimate	Number moving > static	Number static > moving	Z statistic
.1Hz-.5Hz	Smooth	Correlated	19	37	2.602*
		Non-correlated	18	38	2.896*
	Intermittent	Correlated	20	36	2.496*
		Non-correlated	22	34	2.300*
.1Hz-1.0Hz	Smooth	Correlated	63	77	1.668
		Non-correlated	51	89	4.010*
	Intermittent	Correlated	65	75	1.579
		Non-correlated	55	85	3.977*

\* Significant at  $p < .05$  using normal approximation.

Table 9.5: The Wilcoxon analyses indicated that power spectral density of the correlated and non-correlated tracking error was reduced using a moving auditory presentation relative to a static auditory presentation.

correlated error power spectral density of the moving auditory presentation mode and the static auditory presentation mode using both target descriptions. A Wilcoxon signed-rank test was also utilized to determine if a statistically significant difference existed between the non-correlated error power spectral density of the moving auditory presentation mode and the static auditory presentation mode using both target descriptions. The results of these four Wilcoxon signed-rank tests are shown in Table 9.5. The Wilcoxon analyses indicated that the power spectral density of the correlated tracking error was reduced using a moving auditory presentation relative to a static auditory presentation in the frequency range from 0.1Hz to 0.5Hz. The Wilcoxon analyses also indicated that the non-correlated tracking error power spectral density was reduced using a moving auditory presentation relative to a static auditory presentation in the frequency range from 0.1Hz to 1.0Hz.

#### Target movement characteristics affecting augmented tracking performance

The target utilized in this experiment moved only in azimuth. The time history of the target azimuth movement is shown in Figure 9.2. The target azimuth movement was generated by the sum of 3 sinusoids with frequencies of 0.14Hz, 0.21Hz, and 0.31Hz, and was sampled at 50 samples per second. In contrast to the apparent smooth target movement obtained from the 50Hz sampling of the three sinusoids, intermittent 400ms periods when the visual target was non-observable were incorporated into the target motion. Characteristics of the target movement around the intermittent periods, such as target speed at period onset, and angular extent of the movement during the period, varied greatly across periods. Several of these characteristics are shown in Table 9.1. The combination of the high sampling rate of the visual target movement and the embedded intermittent periods provided differing levels of visual target motion predictability. The effect of the target movement characteristics on the difference in tracking performance between the two auditory presentation conditions can be determined through a linear regression analysis.

Variable name	Description	Units
$G$	Extent of target movement during period	degrees
$A_s$	Acceleration of target at period onset	degrees-sec <sup>-2</sup>
$V_s$	Velocity of target at period onset	degrees-sec <sup>-1</sup>
$V_e$	Velocity of target at end of period	degrees-sec <sup>-1</sup>
$V_{s-e}$	Velocity difference between period onset and end	degrees-sec <sup>-1</sup>
$E_{50}$	Tracking error difference during 1 second interval following period	degrees
$E_{20}$	Tracking error difference during 400ms interval following period	degrees
$E_{gap}$	Tracking error difference during period	degrees

Table 9.6: Characteristics of the target movement associated with the intermittent periods, shown in Table 9.1, were used to derive the independent variables used in the regression analysis. The tracking error terms utilized as dependent variables were obtained by subtracting the r.m.s. tracking error obtained under the static auditory presentation mode from the r.m.s. tracking error obtained under the moving auditory presentation mode.

The results of the analyses of variance, shown in Table 9.2, Table 9.3, and Table 9.4, indicated that two large contributors to variance in these analyses were period and subject. In this linear regression analysis, tracking errors were averaged across subjects yielding a single tracking error term for each of the twenty-four periods under each of the two auditory presentation modes. However, several of the periods occurred when the target was extremely close to the left or right edge of the visual display which may have allowed the subject to modify their tracking strategy. To eliminate this potential from the regression analysis, the intermittent periods in which the target angular extent was greater than  $\pm 13^\circ$  were discarded from the analysis. The figure of  $13^\circ$  was chosen arbitrarily. Specifically, the five periods discarded from the regression analysis were period four, eleven, twelve, fourteen, and twenty-one. Thus, nineteen periods remained in the analysis and at each of the nineteen periods, two auditory presentation modes were represented.

Three tracking error terms were utilized as individual dependent variables. These tracking error terms were obtained by subtracting the r.m.s. tracking error obtained under the static auditory presentation mode from the r.m.s. tracking error obtained under the moving auditory presentation mode associated with each intermittent period. The r.m.s. tracking errors were calculated for the intervals during each intermittent period, for the 400ms immediately following each intermittent period, and for the one second interval immediately following each intermittent period. The independent variables used in the regression analysis were derived from the target movement characteristics from Table 9.1. The independent and dependent variables using in the regression analysis are described in Table 9.6.

Three regression equations were analyzed. These three equations are shown below. The absolute value of the velocity and acceleration was taken before the analyses to eliminate bias introduced into the regression analyses due to the direction of motion.



Regression analysis results for $E_{gap}$ , $r^2 = .36$						
Variable	Coefficient	Standard Error	Standard Coefficient	Tolerance	$t$	$p$
$\beta_0$	-0.128	0.136	0.00	.	-0.940	0.364
$G$	-0.118	0.114	-1.886	0.015	-1.034	0.320
$A_s$	-0.014	0.015	-0.727	0.080	-0.926	0.371
$V_e$	-0.037	0.033	1.348	0.034	1.119	0.283
$V_s$	0.016	0.031	0.536	0.049	0.537	0.601
$V_{s-e}$	0.047	0.048	0.789	0.076	0.983	0.344

Source	SS	df	MS	F
Regression	0.189	5	0.038	1.475
Residual	0.334	13	0.026	

Table 9.7: A linear regression analysis was performed on the tracking error difference between trials under the moving auditory presentation and trials under the static auditory presentation averaged across subjects. The tracking error utilized in this analysis was obtained from a one second interval beginning at the end of each intermittent period.

$$E_{gap} = \beta_1 G + \beta_2 |V_s| + \beta_3 |V_e| + \beta_4 |V_{s-e}| + \beta_5 |A_s| + \beta_0 \quad (9.3)$$

$$E_{20} = \beta_1 G + \beta_2 |V_s| + \beta_3 |V_e| + \beta_4 |V_{s-e}| + \beta_5 |A_s| + \beta_0 \quad (9.4)$$

$$E_{50} = \beta_1 G + \beta_2 |V_s| + \beta_3 |V_e| + \beta_4 |V_{s-e}| + \beta_5 |A_s| + \beta_0 \quad (9.5)$$

The results of the linear regression analysis are summarized in Tables 9.8, 9.9, and 9.7. The results of the linear regression analysis indicated that the regression model was not significant for  $E_{20}$ , and  $E_{gap}$  but was significant for  $E_{50}$  ( $F(5, 13) = 3.16, p < 0.05$ ). The  $E_{50}$  linear regression analysis indicated a significant relationship between  $|V_{s-e}|$  and  $E_{50}$  with a coefficient of 0.124 ( $t(18) = 2.778, p < 0.05$ ). The linear regression analysis of  $E_{50}$  also indicated a significant  $\beta_0$  of -0.421. The  $r^2$  resulting the  $E_{50}$  linear regression was 0.55.

Of the three regression equations analyzed, only the equation relating  $E_{50}$  to the target movement characteristics was a statistically significant model. Within this model, only one of the target movement characteristics,  $|V_{s-e}|$ , significantly affected  $E_{50}$ . A secondary model relating  $E_{50}$  to  $|V_{s-e}|$  was constructed and is shown below.

$$E_{50} = \beta_1 |V_{s-e}| + \beta_0 \quad (9.6)$$

A linear regression analysis was performed using the secondary model. The results of this regression analysis are summarized below.

Regression analysis results for $E_{20}$ , $r^2 = .40$						
Variable	Coefficient	Standard Error	Standard Coefficient	Tolerance	$t$	$p$
$\beta_0$	-0.452	0.195	0.00	.	-2.323	0.037*
$G$	0.070	0.163	-0.758	0.015	-0.430	0.674
$A_s$	-0.016	0.021	-0.591	0.080	-0.778	0.451
$V_e$	0.018	0.047	0.444	0.034	0.381	0.710
$V_s$	0.024	0.044	0.525	0.049	0.543	0.596
$V_{s-e}$	0.082	0.068	0.939	0.076	1.209	0.248
Source		SS	df	MS	F	
Regression		0.461	5	0.092	1.749	
Residual		0.686	13	0.053		

\* Significant  $p < .05$

Table 9.8: A linear regression analysis was performed on the tracking error difference between trials under the moving auditory presentation and trials under the static auditory presentation averaged across subjects. The tracking error utilized in this analysis was obtained from a 400ms interval beginning at the end of each intermittent period.

Regression analysis results for $E_{50}$ , $r^2 = .55$						
Variable	Coefficient	Standard Error	Standard Coefficient	Tolerance	$t$	$p$
$\beta_0$	-0.421	0.128	0.00	.	-3.301	0.006*
$G$	0.150	0.107	2.149	0.015	1.401	0.185
$A_s$	-0.024	0.014	-1.155	0.080	-1.749	0.104
$V_e$	-0.026	0.031	-0.860	0.034	-0.849	0.411
$V_s$	-0.032	0.029	-0.944	0.049	-1.122	0.282
$V_{s-e}$	0.124	0.045	1.876	0.076	2.778	0.016*
Source		SS	df	MS	F	
Regression		0.358	5	0.072	3.159*	
Residual		0.295	13	0.023		

\* Significant  $p < .05$

Table 9.9: A linear regression analysis was performed on the tracking error difference between trials under the moving auditory presentation and trials under the static auditory presentation averaged across subjects. The tracking error utilized in this analysis was obtained from a one second interval beginning at the end of each intermittent period.

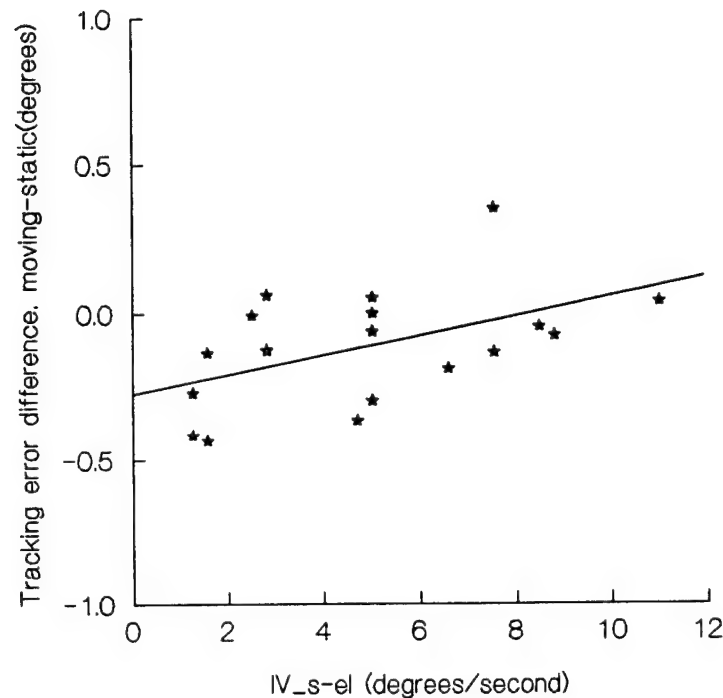


Figure 9.17: The influence of the auditory presentation had a significant relationship to the difference between the velocity of the target at the onset of each intermittent period and the velocity of the target at the end of each intermittent period.

Regression analysis results for $E_{50}$ , $r^2 = .23$						
Variable	Coefficient	Standard Error	Standard Coefficient	Tolerance	$t$	$p$
$\beta_0$	-0.263	0.079	0.00	.	-3.353	0.004*
$V_{s-e}$	0.031	0.014	0.475	1.0	2.227	0.040*
Source		SS	df	MS	F	
Regression		0.147	1	0.147	4.959*	
Residual		0.505	17	0.030		

\* Significant  $p < .05$

Table 9.10: A secondary linear regression analysis was performed relating the tracking error difference between trials under the moving auditory presentation and trials under the static auditory presentation averaged across subjects. The tracking error utilized in this analysis was obtained from a one second interval beginning at the end of each intermittent period. The model included only one target movement characteristic,  $|V_{s-e}|$ .

The results from the linear regression analysis of the secondary model indicated that  $|V_{s-e}|$  significantly affected  $E_{50}$ . A plot of the relationship between  $|V_{s-e}|$  and  $E_{50}$  is shown in Figure 9.17.

### Summary of results and discussions for experiment eight

Seven subjects were run in experiment eight to evaluate the performance implications in an intermittent tracking task in which the moving target to be tracked was either visual or visual and auditory in nature. The intermittent tracking task was a manual control pursuit tracking task in which the visual target sporadically became non-observable for 400ms periods. During these periods, the auditory target remained latched at the last observable position of the visual target. There were 24 intermittent periods included in each 144 second trial.

The overall tracking performance of each trial was not significantly affected by the presence of moving auditory stimuli. However, the tracking error within the intermittent periods as well as within a 400ms interval following the end of each intermittent period was significantly affected by characteristics of the intermittent period.

There was both correlated and non-correlated tracking error in the time histories from the moving and static auditory presentation modes. It was found that there was significant reduction in correlated and non-correlated tracking error power spectral density attributable to the inclusion of auditory target movement. This reduction occurred in a frequency band of 0.1Hz to 0.5Hz for the correlated error and 0.1Hz to 1.0Hz for the non-correlated error.

A linear regression analysis was performed on the tracking error difference between trials under the moving auditory presentation and trials under the static auditory presentation averaged across subjects. The tracking error obtained within the intermittent periods, during the 400ms interval following the intermittent periods, and 1 second following the intermittent periods were utilized as three dependent variables for the analysis. The dependent variables utilized were dynamic characteristics of the target movement associated with each intermittent period, such as the target velocity at the onset of the period. The regression analyses indicated that the absolute value of the difference between the target velocity at the onset of the period and the velocity of the target at the end of the period significantly affected the tracking error difference during the 1 second interval following the period.

It is concluded from this experiment that the small *inter*-sensory influence seen in the previous seven experiments did affect the pursuit tracking of the intermittent target. It is possible that the *intra*-sensory target movement characteristics associated with the intermittent periods, which introduced frequencies higher than the three fundamental sinusoids of the target movement, and the high stroboscopic rate outside of the intermittent periods, may have masked the small auditory influence effect in the analysis of variance of the overall tracking error. This possibility is drawn from Figures 9.13 and 9.15 which depict a reduction in

correlated tracking error within the frequency band of 0.1Hz to 0.5Hz, the results of the Wilcoxon signed-rank tests, and the existence of correlated and non-correlated error above 1.0Hz which does not appear to exhibit the reduction in tracking error. This conclusion is also supported by the regression analysis which indicates that the tracking performance difference observed between the auditory presentation modes during the 1 second interval following the intermittent period is a function of a dynamic characteristic of the target movement.

## Chapter 10

# Conclusions, discussions, and recommendations

### 10.1 Conclusions

The understanding of human motion perception has broad application and fundamental importance. This report increases knowledge of human motion perception by investigating the influence of moving auditory stimuli on visual apparent motion perception. Perceptual and physiological *intra*-sensory and *inter*-sensory literature regarding visual and auditory apparent motion perception was reviewed. Perceptual organization of visual stimuli, driven by inter-stimulus interval (ISI) and angular extent, was measured in the presence of moving and non-moving auditory stimuli. Performance in a manual tracking task using a visual and auditory target was also investigated.

This report provides evidence that visual apparent motion perception can be affected by moving auditory stimuli. This was established within this report through converging empirical results obtained using several experimental techniques.

The overall conclusions of this experimental work are organized into the four following sub-sections. These sub-sections describe overall conclusions from the experimental work and summarize the support for these conclusions.

#### 10.1.1 Moving auditory stimuli can influence angular-extent-driven visual perceptual organization

The support for the conclusion that moving auditory stimuli can influence angular-extent-driven visual perceptual organization comes from results of experiment five. In that experiment, two perceptual organizations could be perceived by the subject, a vertical motion organization or a horizontal motion organization. The addition of a moving auditory source linked spatially and temporally with the horizontal organization of the visual stimuli significantly decreased the number of

vertical motion reports. This is graphically depicted in Figure 7.11.

Two estimates of the strength of this influence can be derived from experiment four and five. The maximum influence exhibited in experiment four was approximately 15%, which is graphically depicted in Figure 7.3 and referenced in Table 7.4. Two parallel lines can be constructed on Figure 7.11, one based on the three means under the moving auditory condition, and one based on the three means under the stationary auditory condition. By cutting these two lines with a horizontal line at the 50% reporting rate, the difference in horizontal separation between the moving auditory condition line and the stationary auditory line represents an estimate of effect strength. In Figure 7.11, at the 50% rate, the stationary auditory line threshold is  $5.06^\circ$  and the moving auditory line threshold is  $6.02^\circ$ . The difference is  $0.96^\circ$  and represents a shift from the stationary auditory threshold of approximately 19%. This estimate is crude at best because only three points are used for each line estimate and the slope of each line is relatively low. However, the strength estimate of 19% compares favorably with the maximum influence exhibited in the time-blocked data from experiment four of 15%.

### **10.1.2 Moving auditory stimuli can influence ISI-driven visual perceptual organization**

The support for the conclusion that moving auditory stimuli can influence ISI-driven visual perceptual organization comes from results of experiment seven. In that experiment, two perceptual organizations could be perceived by the subject viewing a spatially alternating 3-dot/3-dot display, an element motion organization or a group motion organization. The addition of a moving auditory source linked with the horizontal movement of the 3-dot/3-dot visual display significantly decreased the number of group motion reports. This is graphically depicted in Figure 8.9. The shift in ISI attributable to the decrease in group motion reports was approximately 6.5% within ISIs ranging from 83ms to 150ms.

### **10.1.3 Temporal factors can mask inter-sensory influence effects on perceptual organization**

There are several indications that temporal factors, such as perceptual hysteresis, manifested as perceptual organizational capture, can mask the small inter-sensory influence effect. These indications are apparent when results from experiment six are compared with results from experiment seven, and when results from experiment four are compared with results from experiment five.

Experiment four utilized trials in which the spatial extent was systematically increased or decreased until a change in perceptual organization was reported by the subject. In this manner, the threshold horizontal extent at which the perceptual organization would breakdown was measured. In this methodology, the subject perceived a single organization over a period of approximately 45 seconds. The mean

threshold horizontal extents reported by the subjects during the ascending trials were smaller than those reported by the subjects during the descending trials. However, when the threshold horizontal extents were viewed as a function of time, there was a significant difference between the results from the initial 4 trials of the experimental session and the overall results. In contrast, experiment five clearly demonstrated a significant influence effect of the auditory stimuli on the perceptual organization under horizontal extents ranging from  $5.22^\circ$  to  $5.76^\circ$  using a vertical extent of  $5.0^\circ$ . The trial length in experiment seven was approximately 7.5 seconds and the horizontal extent utilized in each stimulus presentation was randomly selected. When the results of experiment five were viewed as functions of time, no significant changes in results were observed during the length of the experiment session.

The experimental technique utilized in experiment five reduced the potential for hysteresis by randomizing the horizontal extents of each stimulus presentation. The potential for hysteresis was also reduced in experiment four by separating each stimulus presentation with a display screen to prompt a report from the subject.

The ascending and descending series presented the subject in experiment four with two linkages between the visual and auditory stimuli when the auditory stimulus was moving. In the case of the ascending series, the visual motion was initially perceived as being horizontal, and the auditory stimulus was also moving horizontally. As the series progresses, that linkage was reinforced until the subject reported vertical motion, at which time the visual and auditory stimuli would have been conflict. In contrast to this, the descending series began in conflict, with strong vertical motion visually and strong auditory horizontal motion. The descending series remained in conflict until the subject reported horizontal motion, at which time the auditory and visual perceptions were again unified. As these series alternated during the experimental session, the subjects may have slowly de-coupled the auditory and visual links to resolve the almost continual inter-sensory conflict, potentially over the course of 5 to 10 minutes.

From the contrasting results in experiment four and five, it is concluded that the effect of linked auditory stimuli on spatially-driven visual motion perception organization can be masked by a temporal effect related to the lack of correlation between the auditory and visual displays over the duration of an experimental session.

The contrast between experiment six and seven indicated a different temporal effect than the contrast between experiment four and five. The contrast between experiment six and seven indicated that perceptual hysteresis, or perceptual capture, could affect measurement of auditory influence on visual apparent motion perception. Experiment six utilized trials in which the ISI was systematically increased or decreased until a change in perceptual organization was reported by the subject. In this manner, the threshold ISI at which the perceptual organization would breakdown was measured. In this methodology, the subject perceived a single organization over a period of approximately 23 seconds. The mean threshold ISIs reported by the subjects during the ascending trials were larger than those reported



by the subjects during the descending trials and thus, indicated the presence of hysteresis. The results from experiment six indicated that there was no significant influence of the auditory stimuli on the threshold ISIs. In contrast, experiment seven clearly demonstrated a significant influence effect of the auditory stimuli on the perceptual organization under ISIs ranging from 83ms to 150ms. The trial length in experiment seven was approximately 5 seconds and the ISI utilized in each stimulus presentation was randomly selected.

The experimental technique utilized in experiment seven reduced the potential for hysteresis by randomizing the ISIs of each stimulus presentation. This was in contrast to experiment six in which the ISI was monotonically adjusted throughout each stimulus presentation. The perceptual organization change measured in experiment six may occur at the point where perceptual hysteresis breaks down. The ISI at which hysteresis may break down may not necessarily be identical to the ISI at which perceptual organization may switch without hysteresis present.

From these results, it is concluded that the effect of linked auditory stimuli on temporally-driven visual motion perception organization can be masked by perceptual hysteresis effects.

#### 10.1.4 Pursuit tracking of a visual target can be affected by auditory target augmentation

It is established within this report that moving auditory stimuli can influence the perceptual organization of visual apparent motion stimuli. This *inter*-sensory influence is small relative to the influence of *intra*-sensory stimulus characteristics and can be masked by perceptual hysteresis in both spatial and temporal domains.

This influence, however, was evaluated utilizing psychophysical techniques requiring simple forced-choice responses to be made by the subjects. These psychophysical experiments were augmented by investigating whether the small auditory influence could affect the performance of a higher-level task. Seven subjects were run in experiment eight to evaluate the performance implications in an intermittent tracking task in which the moving target to be tracked was either visual or visual and auditory in nature. The intermittent tracking task was a manual control pursuit tracking task in which the visual target sporadically became non-observable for 400ms periods. During these periods, the auditory target remained latched at the last observable position of the visual target. There were 24 intermittent periods included in each 144 second trial.

Tracking performance was significantly affected by the presence of moving auditory stimuli. This effect was not apparent in time domain analyses of the tracking error but became apparent in frequency domain analyses as well as linear regression analyses of target movement characteristics.

There was both correlated and non-correlated tracking error in the time histories from the moving and static auditory presentation modes. It was found that there

was significant reduction in correlated and non-correlated tracking error power spectral density attributable to the inclusion of auditory target movement. This reduction occurred in a frequency band of 0.1Hz to 0.5Hz for the correlated error and 0.1Hz to 1.0Hz for the non-correlated error.

Linear regression analyses were performed on the tracking error difference, averaged across subjects, between trials under the moving auditory presentation and trials under the static auditory presentation. The tracking errors obtained within the intermittent periods, during the 400ms interval following the intermittent periods, and 1 second following the intermittent periods, were utilized as three dependent variables for the analyses. The independent variables utilized were dynamic characteristics of the target movement associated with each intermittent period, such as the target velocity at the onset of the period. The regression analyses indicated that the absolute value of the difference between the target velocity at the onset of the period and the velocity of the target at the end of the period significantly affected the tracking error difference during the 1 second interval following the period.

It is concluded from this experiment that the small *inter*-sensory influence seen in the previous seven experiments did affect the pursuit tracking of the intermittent target. It is possible that the *intra*-sensory target movement characteristics associated with the intermittent periods, which introduced frequencies higher than the three fundamental sinusoids of the target movement, and the high stroboscopic rate outside of the intermittent periods, may have masked the small auditory influence effect in the analysis of variance of the overall tracking error. This possibility is drawn from Figures 9.13 and 9.15 which depict a reduction in correlated tracking error within the frequency band of 0.1Hz to 0.5Hz, the results of the Wilcoxon signed-rank tests, and the existence of correlated and non-correlated error above 1.0Hz which does not appear to exhibit the reduction in tracking error. This conclusion is also supported by the regression analysis which indicated that the tracking performance difference observed between the auditory presentation modes during the 1 second interval following the intermittent period is a function of a dynamic characteristic of the target movement.

## 10.2 Summary of conclusions

Eight experiments were performed and described within this report. Perceptual organization of visual stimuli, driven by inter-stimulus interval (ISI) and angular extent, was measured in the presence of moving and non-moving auditory stimuli. Performance of a manual tracking task using a combined visual-auditory target was also investigated. Small effects of contemporaneous moving auditory stimuli on angular-extent-driven and ISI-driven visual perceptual organizations were measured in five experiments. The effects were susceptible to perceptual hysteresis and existed only when visual-based organization was ambiguous. When angular-extent-driven stroboscopic visual stimuli were augmented with moving auditory stimuli, a maximum shift of 15% to 19% in the angular extent of stroboscopic visual stimuli

required to maintain perceptual organization was measured. Additionally, a shift of 6.5% in the ISI required to maintain perceptual organization was measured, within ISIs ranging from 83ms to 150ms, when stroboscopic visual stimuli were augmented with moving auditory stimuli.

Dynamic characteristics of an auditory localizer, which was used in a large portion of the work, were evaluated to assess the generalized nature of the experimental results. The auditory localizer generated auditory stimuli that elicited velocity discriminations of 14% using velocities ranging from  $20^{\circ} \text{ sec}^{-1}$  to  $100^{\circ} \text{ sec}^{-1}$ . The minimum auditory movement angle measured using the auditory localizer was  $8.1^{\circ}$  at  $90^{\circ} \text{ sec}^{-1}$ , which differed from previous studies in the literature using real stimuli by less than 3%.

Tracking of an intermittent visual-auditory target, relative to tracking of an intermittent visual-only target, was affected by characteristics of the target movement and a reduction in the power spectral density of correlated and non-correlated tracking error was observed between 0.1Hz to 0.5Hz.

In summary, this report is consistent with the premise that visual apparent motion perception can be influenced by moving auditory stimuli and that manual tracking performance can be affected by the presence of a linked auditory and visual target.

### 10.3 Implications of conclusions on the inter-modal process model

The combined empirical results of experiment one, four, five, six, and seven provide a foundation for modification of the inter-modal model depicted in Figure 3.25. The combined empirical results provide evidence that there is a small auditory influence of visual apparent motion perception. This is a new finding in that it is not documented within the literature. This new finding supports the premise that inter-modal connection exists between the visual and auditory motion perception systems.

In addition to the existence of the inter-modal influence, several characteristics of this influence are derivable from the results of experiment one, four, five, six and seven. These characteristics are described in the following paragraphs.

It is clear that when movement of auditory stimuli is approximately equal to, or lower than, the dynamic resolution of the auditory system, the influence of the stroboscopic auditory stimuli on visual apparent motion perception is non-existent or severely diminished. Thus, it can be concluded that, in a coarse sense, the strength of the influence is indeed a function of the perceptual strength elicited by the auditory stimulus. To reflect this, the model in Figure 3.25 must be modified to accommodate an influence that is a function of the auditory perceptual strength. However, the function relating influence strength to auditory perceptual strength is not known.

The entrance location of the auditory influence within the visual apparent motion

perception model in Figure 3.25 is suggested by contrasts between results from experiment one, four, five, six, and seven. Several locations within the inter-modal process model could support processing to achieve the influence of stroboscopic auditory stimuli on visual apparent motion perception found in the results of experiment one, four, and five. Within experiment one, four and five, the independent variable which was systematically manipulated was the extent of the horizontally-orientated stroboscopic display. The influence effect could have been a result of an influence in either the ISI processor, the extent processor, the comparator supporting the perceptual organization, or within the decision processor as a criterion modification. The location of the auditory influence processing within the inter-modal process model can be narrowed by examining characteristics of the inter-modal influence.

In experiment five, the influence of the auditory stimuli increased the spatial separation of the visual stimuli at which horizontal motion would be judged stronger than the vertical motion, but only over a relatively small range of horizontal separation distances. Outside this small range, there was no influence. Within the small range of separation distances, the percentage of vertical to horizontal motion reports ranged from approximately 35% to 50%. This range can be considered an ambiguous perceptual area. For these reasons, one likely candidate for model modification is the decision criteria within the decision processor. Further support for decision processor implication comes from results of experiment seven. It was hypothesized in experiment seven that the inter-modal influence occurred due to an enhancement in the strength of visual apparent motion brought about by the contemporaneous presentation of moving auditory stimuli. In experiment seven, the presence of this enhancement would not have affected the perceptual organization of the visual stimuli based upon architecture of the mechanism model and process model. However, over a range of ISIs from 83ms to 150ms in experiment seven, the auditory influence was present as reflected by an increase in the percentage of element motion reports over group motion reports. The percentage of group to element motion reports in this experiment ranged from approximately 15% to 70%. The existence of the influence using this stimulus provides additional supporting evidence that the location of the auditory influence introduction into the visual motion perception processing falls beyond the comparator processing and may be located within, or beyond, the decision processor.

Dissipation of the auditory influence on visual apparent motion was found to occur within results of experiment four but not experiment five. This contrast in results indicates that conflicts between the direction of perceived visual motion and the direction of perceived auditory motion may have caused the dissipation. The use of ascending and descending trials in experiment four provided the subjects with long exposures to non-linked visual and auditory stimuli. However, experiment five did not present the subject with long duration non-linked visual and auditory stimuli. In experiment four, the influence of the stroboscopic auditory stimuli on visual apparent motion perception was present in the initial trials of the experimental period but dissipated over time. In contrast, no dissipation was found in results from

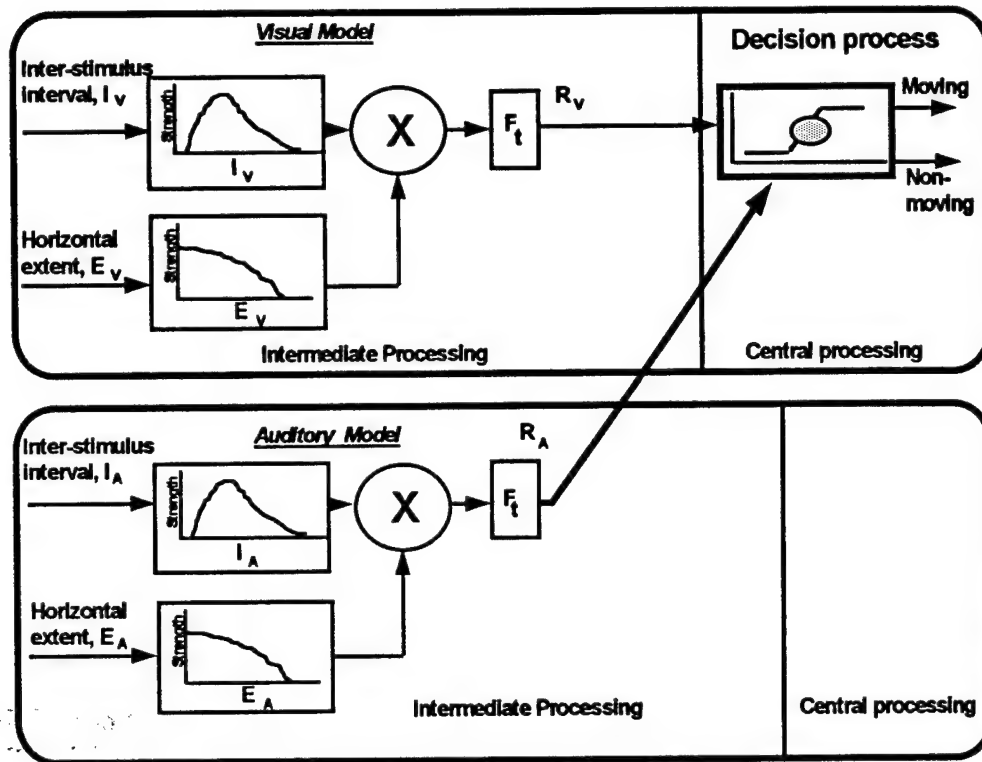


Figure 10.1: The model depicted in Figure 3.25, modified to accommodate the implication of the experimental results described within this report, is depicted in this Figure.

experiment five. The dissipation resulting from visual-auditory stimulus non-linkage indicates that a correlation between the visual and auditory motion perceptions may be present. To accommodate this characteristic, correlation processing between the visual and auditory motion perceptual system must be added to the inter-modal processing model. Characteristics of the correlation processing between the visual and auditory perceptual system were not systematically investigated within these experiments and thus, specific characteristics can not be determined from the empirical data. However, it does appear that the effect of non-correlation between the visual and auditory stimuli takes several minutes to manifest itself upon the auditory influence on visual motion perception.

The model depicted in Figure 3.25, modified to accommodate the implication of the experimental results described within this report, is depicted in Figure 10.1.

## 10.4 Recommendations

This report supports the premise that visual apparent motion perception can be minimally influenced by moving auditory stimuli. Results of the literature search and experimental work within this report illuminate several potential areas of follow-on research. These areas are described in the following sections.

### 10.4.1 Visual and auditory linkage characteristics

The empirical work described within this report typically used stroboscopic stimuli having two conditions of temporal and spatial linkage, those conditions being linked or non-linked. One additional issue not addressed by this approach is how the auditory influence may be affected by the temporal and spatial correlation of inter-sensory stimuli. Determining the function relating influence to inter-sensory correlation would further illuminate the processing depicted in Figure 3.24 between  $R_v$  and  $R'_v$ .

Experiments could be performed to bound the temporal and spatial properties of the auditory display that allow perceptual linkage between visual and auditory stimuli to be established and maintained by the observer. These experiments could use a single dot visual display that would be continually moving. A single auditory source would track the moving visual target with varying amounts of temporal lag. The speed of the visual target could also be varied in a controlled fashion such that the temporal lag would introduce spatial lag that is proportional to the visual target speed. The observer could be asked to report the state of the linkage. Data could then be analyzed to determine what temporal and spatial properties were needed to establish initial linkage and what temporal and spatial properties were needed to dissipate linkage.

### 10.4.2 Aiding wide off-boresight tracking with auditory displays

The experiments described within this report were limited to small spatial extents relative to the field-of-regard of both the auditory and visual systems. By expanding the spatial extent of the stimuli to include the full auditory and visual field-of-regard, a broader fundamental understanding of the effect of spatial and temporal extent on visual and auditory interactions at the perceptual and cognitive levels might be achieved. This understanding could be a basis for developing wide field-of-view multi-sensory displays.

Research could be conducted to investigate the augmentation of the visual display of information in the forward hemisphere with auditory display of information in the rearward hemisphere. As an example of this type of display, if a target was in front of the subject, it might be presented visually, as the target traveled from the front to the rear of the subject, the display might go from visual to auditory, perhaps with some period of overlapping displays. The subject's ability to maintain awareness of

the spatial relationship of a large number of targets may be enhanced utilizing a display technique based on the characteristics of the human's ability to combine inter-sensory across large fields-of-view.

#### 10.4.3 Tracking of non-linked visual-auditory targets

The presentation of auditory information other than position may affect the performance of complex visual tasks differently than the presentation of auditory stimuli that are linked in position with visual stimuli. By altering the level of cognitive processing required to interpret the auditory display of information, the understanding of how cognition may affect inter-sensory influence may be enhanced. In this way, the relative contributions of perception and cognition may be investigated within complex task environments.

A series of experiments could be conducted using the auditory display to indicate the error between a visual target and a visual cursor in a tracking task. The gain of the error displayed aurally could be used as an independent variable.

#### 10.4.4 Interaction between intermittent, multi-sensory target movement and tracking performance

The effect of the auditory display within the tracking task used in experiment eight was affected by the difference of target velocity at the onset and end of the intermittent periods. It may be possible to isolate other target movement characteristics which contribute to the influence effect and under what conditions these contributions may exist. A series of experiments could be performed which investigated the implications of various target movement patterns prior to the intermittent periods in a pursuit or compensatory tracking task using both auditory and visual targets. The target movement patterns could be classified by their potential predictability. Correlation between predictability and tracking performance could be established. The effect of the presence of the moving auditory display could be assessed relative to predictability of the target movement.



# Bibliography

- [1] E. H. Adelson and J. R. Bergen. Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America A*, 2(2):284-299, 1985.
- [2] J.A. Altman and O.V. Viskov. Discrimination of perceived movement velocity for fused auditory image in dichotic stimulation. *Journal of the Acoustical Society of America*, 61:816-819, 1977.
- [3] Stuart Anstis and V. S. Ramachadrab. Visual inertia in apparent motion. *Vision Research*, 27(5):755 - 764, 1987.
- [4] C. Auerbach and P. Sperling. A common auditory-visual space: Evidence for its reality. *Perception and Psychophysics*, 16:129-135, 1974.
- [5] Norman I. Badler and John K. Tsotsos, editors. *Motion: Representation and Perception*. North Holland, New York, 1986.
- [6] Kartene Ball, Robert Sekuler, and John MacHamer. Detection and identification of moving targets. Technical Report AFOSR-TR-85-0610, Northwestern University, June 1985.
- [7] Alfred Barbour. An investigation of dual sensory stimulus presentation of complex noise-like sounds. Technical Report TM-79-115 (AD-A074936), Pennsylvania State University, June 1979.
- [8] Robert J. Baron. *The Cerebral Computer, An Introduction to the Computational Structure of the Human Brain*. Lawrence Erlbaum, New Jersey/London, 1987.
- [9] Brent Baxter. *Image Processing in the Human Visual System*. PhD thesis, University of Utah, December 1975. UTEC-CSc-75-168.
- [10] Julius S. Bendat and Allan G. Piersol. *Engineering Applications of Correlation and Spectral Analysis*. John Wiley & Sons, New York, New York, 1980.
- [11] Jens Blauert. *Spatial Hearing*. The MIT Press, Cambridge, Massachusetts, 1983.
- [12] Kenneth R. Boff, Lloyd Kauffman, and James P. Thomas, editors. *Handbook of Perception and Human Performance*. John Wiley, New York, 1986.



- [13] Kenneth R. Boff and Janet E. Lincoln, editors. *Engineering Data Compendium - Human Perception and Performance*. Armstrong Aerospace Medical Research Laboratory, Wright Patterson Air Force Base, Ohio, 1988.
- [14] O. Braddick. The masking of apparent motion in random-dot patterns. *Vision Research*, 13:355-369, 1973.
- [15] Bruno Breitmeyer, James G. May, and Mary C. Williams. Spatial frequency and contrast effects on percepts of bistable stroboscopic motion. *Perception and Psychophysics*, 44(6):525-531, 1988.
- [16] Bruno G Breitmeyer and Alysia Ritter. Visual persistence and the effect of eccentric viewing, element size, and frame duration on bistable stroboscopic motion percepts. *Perception and Psychophysics*, 39(4):275-280, 1986.
- [17] Peter Burt and George Sperling. Time, distance, and feature trade-offs in visual motion perception. *Psychological Review*, 88(2):171-195, 1981.
- [18] Neil R. Carlson. *Foundations of Physiological Psychology*. Allyn and Bacon, Needham Heights, Massachusetts, 1992.
- [19] Leo M. Chalupa and Robert W. Rhoades. Responses of visual, somatosensory, and auditory neurones in the golden hamster's superior colliculus. *Journal of Physiology*, 270:595-626, 1977.
- [20] James Cutting and Jeffrey J. Garvin. Fractal curves and complexity. *Perception and Psychophysics*, 42(4):365-370, 1987.
- [21] James P. Dixon. Single unit recordings of cells responsive to visual, somatic, acoustic, and noxious stimuli in the superior colliculus of the golden hamster. Technical Report ADA091548, Virginia Commonwealth University, August 1978.
- [22] Theodore J. Doll. Development of simulated directional audio for cockpit applications. Technical Report AAMRL-TR-86-014, Georgia Institute of Technology, January 1986.
- [23] A. J. van Doorn and J. J. Koenderink. Temporal properties of the visual detectability of moving spatial white noise. *Experimental Brain Research*, 45:179-188, 1982.
- [24] A. J. van Doorn and J. J. Koenderink. Spatiotemporal integration in the detection of coherent motion. *Vision Research*, 24(1):47-53, 1984.
- [25] Richard Durbin, Christopher Maill, and Graeme Mitchison, editors. *The Computing Neuron*. Addison-Wesley Publishing Company, Wokingham, England, 1989.
- [26] R A Earnshaw, M A Gigante, and H Jones, editors. *Virtual Reality Systems*. Academic Press, Harcourt Brace and Company, San Diego, CA and London, England, 1994.

- [27] Robert G. Eggleston. Apparent motion and prior correspondence effects in visual perception. Technical Report AAMRL-TR-86-027, Armstrong Aerospace Medical Research Laboratory, June 1986.
- [28] Mark A. Erikson and Richard L. McKinley. Auditory localization cue synthesis and human performance. In *Proceedings of the IEEE 1989 National Aerospace Electronics Conference - NAECON 89*, Dayton, Ohio, May 1989. Institute of Electrical and Electronics Engineers.
- [29] Olivier Dominique Faugeras. *Digital Color Image Processing and Psychophysics within the Framework of a Human Visual Model*. PhD thesis, University of Utah, June 1976. UTEC-CSc-77-029.
- [30] David Finlay and Michael Von-Grünau. Some experiments of the breakdown effect in apparent motion. *Perception and Psychophysics*, 42(6):526-534, 1987.
- [31] James D. Foley. Interfaces for advanced computing. *Scientific American*, pages 127-135, October 1987.
- [32] Thomas A. Furness. The super cockpit and its human factors challenges. *Proceedings of the Human Factors Society 30th Annual Meeting*, pages 48-52, 1986.
- [33] Thomas A. Furness. Harnessing virtual space. *Society for Information Display International Symposium*, 1988.
- [34] R.W. Gatehouse. *Localization of Sound*. Amphora Press, Groton, Connecticut, 1982.
- [35] Stan C. A. M. Gielen, Richard A. Schmidt, and Pieter J. M. Van Den Heuvel. On the nature of intersensory facilitation of reaction time. *Perception and Psychophysics*, 34(2):161-168, 1983.
- [36] E. Bruce Goldstein. *Sensation and Perception*. Wadsworth, Belmont California, 1984.
- [37] Barbara Gordon. Receptive fields in deep layers of cat superior colliculus. *Journal of Neurophysiology*, 36:157-178, 1973.
- [38] Norma van Surdam. Graham. *Visual Pattern Analyzers*, volume 16. Oxford University Press, New York, 1989.
- [39] D. Wesley Grantham. Detection and discrimination of simulated motion of auditory targets in the horizontal plane. *Journal of the Acoustic Society of America*, 79(6):1939-1949, 1986.
- [40] D. Wesley Grantham. Motion aftereffects with horizontally moving sound sources in the free field. *Perception and Psychophysics*, 45(2):129-136, 1989.
- [41] D. Wesley Grantham. Adaptation to auditory motion in the horizontal plane: Effect of prior exposure to motion on motion detectability. *Perception and Psychophysics*, 52(2):144-150, 1992.

- [42] David M. Green. Complex auditory signals. Technical Report AFOSR-TR-88-0985 (AD-A199832), University of Florida, October 1988.
- [43] Stephen Grossberg. Cortical dynamics of three-dimensional form, color, and brightness perception, ii: Binocular theory. *Perception and Psychophysics*, 41:117–158, 1987.
- [44] Stephen Grossberg and Ennio Mingolla. Neural dynamics of motion perception: Direction fields, apertures, and resonant grouping. *Perception and Psychophysics*, 53(3):243–278, 1993.
- [45] Stephen Grossberg and Michael E. Rudd. A neural architecture for visual motion perception: group and element apparent motion. *Neural Networks*, 2:421–450, 1989.
- [46] Michael W. Haas. Visually-coupled systems and simulation devices. *Journal of Aircraft*, 21(8):639–640, August 1984.
- [47] Donald J. Harris. *Some Relations between Vision and Audition*. Charles C. Thomas, Springfield, Illinois, 1950.
- [48] J. Donald Harris. A florilegium of experiments on directional hearing. *Acta Oto-Laryngologica*, Supplement 298, 1972.
- [49] J. Donald Harris and R.L. Sergeant. Monaural/binaural minimum audible angle for a moving sound source. *Journal of Speech and Hearing Research*, 14:618–629, 1971.
- [50] Richard Held, Hershel W. Leibowitz, and Hans-Lukas Teuber, editors. *Handbook of Sensory Physiology, Perception*, volume 8. Springer-Verlag, Berlin, 1978.
- [51] Maurice Hershenson. Visual system responds to rotational and size-change components of complex proximal motion patterns. *Perception and Psychophysics*, 42(1):60–64, 1987.
- [52] E.C. Hildreth and C. Koch. The analysis of visual motion: From computational theory to neural mechanisms. *A.I. Memo No. 919*, December 1986.
- [53] Ellen C. Hildreth and Shimon Ullman. The measurement of visual motion. *AI Memo No. 699*, December 1982.
- [54] Ellen Catherine Hildreth. *The Measurement of Visual Motion*. The MIT Press, Cambridge, Massachusetts and London, England, 1984.
- [55] Len Hindus. Virtual reality offers growing opportunity for risk-takers. *EDN, News Edition*, 35(10A):51–60, MAY 1990.
- [56] Berthold KP Horn and Brian G. Schunck. Determining optic flow. *Artificial Intelligence*, 17:185–203, 1981.

- [57] Timothy L. Hubbard and Jamshed J. Bharucha. Judged displacement in apparent vertical and horizontal motion. *Perception and Psychophysics*, 44(3):211-221, 1988.
- [58] Radu S. Jasinchi. Space-time filtering, sampling and motion uncertainty. Technical Report CMU-RI-TR-88-9 (AD-A199667), Carnegie Mellon University, September 1988.
- [59] B. Jones and B. Kabanoff. Eye movements in auditory space perception. *Perception and Psychophysics*, 17:241-245, 1975.
- [60] Mary K. Kaiser. Angular velocity discrimination. *Perception and Psychophysics*, 47:149-156, 1990.
- [61] James Thomas Kajiya. *Toward a Mathematical Theory of Perception*. PhD thesis, University of Utah, Department of Computer Science, December 1978.
- [62] Lloyd Kaufman and Samuel J. Williamson. The perception of the higher derivatives of visual motion. Technical Report AFOSR-TR-87-0435, New York University, March 1987.
- [63] J.W. Kling and Lorrin A. Riggs, editors. *Experimental Psychology*, volume 1. Holt, Rinehart, and Winston, New York, 1972.
- [64] E. I. Knudsen. Synthesis of a neural map of auditory space in the owl. *Journal of Neuroscience*, 2:1177-1194, 1984.
- [65] E. I. Knudsen and M. Konishi. Center-surround organization of auditory receptive fields in the owl. *Science*, 202:778-780, 1978.
- [66] E. I. Knudsen and M. Konishi. Space and frequency are represented separately in auditory midbrain of the owl. *Journal of Neurophysiology*, 41:870-883, 1978.
- [67] David A. Kobus and Lawrence Lewandowski. Bimodal information processing in a sonar task. Technical Report 1075 (AD-A170991), Naval Submarine Medical Research Laboratory, June 1986.
- [68] Gina Kolata. Esoteric math has practical result. *Science*, 225:494-495, August 1984.
- [69] Paul A. Kolers. *Aspects of Motion Perception*, volume 16. Pergamon Press Ltd., Headington Hill Hall, Oxford, 1972.
- [70] Masakazu. Konishi. Listening with two ears. *Scientific American*, April:66-73, 1993.
- [71] Myron W. Krueger. *Artificial Reality II*. Addison-Wesley, Reading, Massachusetts, 1991.
- [72] Thomas K. Kuyk. Using depth recovery in humans. Technical Report AAMRL-TR-88-035, Armstrong Aerospace Medical Research Laboratory, July 1988.

- [73] Stephen Lakatos. Temporal constraints on apparent motion in auditory space. *Perception and Psychophysics*, 54(2):139–144, 1993.
- [74] Nikos K. Logothetis and Jeffrey D. Schall. Neuronal correlates of subjective visual perception. *Science*, 245:761–763, August 1989.
- [75] Ivan D. London. Research on sensory interaction in the soviet union. *Psychological Bulletin*, 51(6):531–568, November 1954.
- [76] Walter Makous, John Maunsell, and Tatiana Pasternak. New insights on visual cortex. Technical Report AFOSR-TR-88-0983 (AD-A199826), University of Rochester, June 1988.
- [77] Lawrence E. Marks. *The Unity of the Senses*. Academic Press, New York, 1978.
- [78] S. Mckee. Visual processing of object velocity and acceleration. Technical Report AFOSR-TR-87-1833, The Smith-Kettlewell Eye Research Foundation, November 1987.
- [79] Suzanne P. McKee. The spatial and temporal parameters of velocity discrimination. Technical Report AFOSR-TR-85-0572 (AD-A158735), The Smith-Kettlewell Institute, December 1984.
- [80] Richard L. McKinley. Concept and design of an auditory localization cue synthesizer. Technical Report AFIT/GE/ENG/88D-29 (AD-A203053), Air Force Institute of Technology, December 1988.
- [81] M. Alex Meridith and Barry E. Stein. Interactions among converging sensory inputs in the superior colliculus. *Science*, 221(4608):389–391, July 1983.
- [82] Jeff Miller. Timecourse of coactivation in bimodal divided attention. *Perception and Psychophysics*, 40(5):331–343, November 1986.
- [83] Brian C. J. Moore. *An Introduction to the Psychology of Hearing*. Harcourt Brace Jovanovich, Academic Press, London, England, 1989.
- [84] Ann O’Leary and Gillian Rhodes. Cross-modal effects on visual and auditory object perception. *Perception and Psychophysics*, 35(6):565–569, 1984.
- [85] A.R. Palmer and A.J. King. Monaural and binaural contributions to an auditory space map in the guinea-pig superior colliculus. In R. Klinke and R. Hartmann, editors, *Hearing-Physiological Bases and Psychophysics*, Bad Nauheim, Germany, April 1983. Springer-Verlag.
- [86] A. J. Pantle and J. T. Petersik. Effects of spatial parameters on the perceptual organization of a bistable motion display. *Perception and Psychophysics*, 27:307–312, 1980.
- [87] Allan Pantle and Lucinda Picciano. A multistable movement display: Evidence for two separate motion systems in human vision. *Science*, 193, August 1976.

- [88] *PC-MATLAB Users Guide*. The MathWorks, Inc., South Natick, Massachusetts, 01760, 1990.
- [89] Alex Pentland. Fractal-based description of natural scenes. Technical Report DTIC AD-P001206, SRI International, August 1983.
- [90] Lauri Pentti. Auditory localization during rotation of the visual environment. Technical Report 2, Psychological Institute, University of Helsinki, 1955.
- [91] David R. Perrott. Auditory psychomotor coordination. *Proceedings from the Human Factors Society, 32nd Annual Meeting*, pages 81–85, 1988.
- [92] David R. Perrott, Vernee Buck, Wendy Waugh, and Thomas Z. Strybel. Dynamic auditory localization: Systematic replication of the auditory velocity function. *The Journal of Auditory Research*, 19:277–285, 1979.
- [93] David R. Perrott and A.D. Musicant. Minimum auditory movement angle: Binaural localization of moving sound sources. *Journal of the Acoustic Society of America*, 62:1463–1466, December 1977.
- [94] David R. Perrott and Juliana Tucker. Minimum auditory movement angle as a function of signal frequency and velocity of the source. *Journal of the Acoustic Society of America*, 83(4):1522–1527, April 1988.
- [95] J. Timothy Petersik. The two-process distinction in apparent motion. *Psychological Bulletin*, 106(1):107–127, 1989.
- [96] J. Timothy Petersik and A. J. Pantle. Factors controlling the competing sensations produced by a bistable stroboscopic display. *Vision Research*, 19:143–154, 1979.
- [97] J. Timothy Petersik and Amy Rosner. The effects of position cues on the appearance of stimulus elements in a bistable apparent movement display. *Perception and Psychophysics*, 48(3):280–284, 1990.
- [98] Alan R. Pinkus. The spatiotemporal characteristics of visual motion priming. Technical Report AL/CF-TR-1994-0084, Armstrong Laboratory, July 1994.
- [99] U.E. Raymond. The interaction of target size and background pattern on perceived velocity during visual tracking. *Perception and Psychophysics*, 43(5):425–430, 1988.
- [100] W.A. Richards, J.M. Rubin, and D.D. Hoffman. Equation counting and the interpretation of sensory data. Technical Report AI Memo 614 (AD-A103924), Massachusetts Institute of Technology, June 1981.
- [101] C.L. Searle, L.D. Braida, M.F. Davis, and H.S. Colburn. Model for auditory localization. *Journal of the Acoustic Society of America*, 60(5):1164–1175, November 1976.

- [102] Robert Sekuler. Enhancing sensitivity to visual motion and enhancing visual sensitivity. Technical Report AFSOR-TR-85-0668, Northwestern University, June 1985.
- [103] Robert Sekuler. Enhancing visual sensitivity. Technical Report AFOSR-TR-85-0683, Northwestern University, May 1985.
- [104] Robert Sekuler and Randolph Blake. *Perception*. Alfred A. Knopf, New York, 1985.
- [105] R.W. Sekuler, D.W. Williams, and S.A. Harp. Sensitivity to visual motion in statistically defined displays. Technical Report ARI Research Note, Northwestern University, Mar 1986.
- [106] Margaret Euphrasia Sereno. *Neural Computation of Pattern Motion*. The MIT Press, Cambridge, Massachusetts, 1993.
- [107] M.L. Shaw. Attention and decision factors in bimodal signal processing. Technical Report AFOSR-TR-82-0712, Rutgers-The State University of New York, August 1982.
- [108] B.R. Shelton and C.L. Searle. The influence of vision on the absolute identification of sound-source position. *Perception and Psychophysics*, 28:589-596, 1980.
- [109] R.D. Sorkin. Auditory heads-up cockpit display systems. Technical Report BRA 87-3, Behavioral Research Associates, October 1987.
- [110] A. R. A. Sovijarvi and H. Hyvarinin. Auditory cortical neurons in the cat sensitive to the direction of sound source movement. *Brain Research*, 73:455-471, 1974.
- [111] Lothar Spillmann and John S. Werner, editors. *Visual Perception, The Neurophysiological Foundations*. Academic Press, San Diego, California, 1990.
- [112] Harold Stanislaw. Methodological considerations for the study of multimodal signal detection. *Perception and Psychophysics*, 44(6):541-550, 1988.
- [113] Barry E. Stein and Mekanjuola O. Arigbede. Unimodal and multimodal response properties of neurons in the cat's superior colliculus. *Experimental Neurology*, 36:179-196, 1972.
- [114] Richard M. Stern and H. Steven Colburn. Theory of binaural interaction based on auditory-nerve data. iv. a model for subjective lateral position. *J. Acoust. Soc. Am.*, 64(1):127-140, JULY 1978.
- [115] Richard M. Stern, Andrew S. Zeiberg, and Constantine Trahiotis. Lateralization of complex binaural stimuli: A weighted-image model. *J. Acoust. Soc. Am.*, 84(1):156-165, JULY 1988.

- [116] T.G. Stockham. Image processing in the context of a visual model. *Proceedings of the IEEE*, 60(7):828-842, July 1972.
- [117] Thomas G. Jr. Stockham. Intra-frame encoding for monochrome images by means of a psychophysical model based on nonlinear filtering of multiplied signals. In *Symposium on Picture Bandwidth Compression*, Massachusetts Institute of Technology, April 1969,. Gordon and Breach, Science Publishers, Inc.
- [118] E.J. Stoffels and M.W. Van Der Molen. Effects of visual and auditory noise on visual choice reaction time in a continuous-flow paradigm. *Perception and Psychophysics*, 44(1):7-14, 1988.
- [119] Thomas Z. Strybel. Perception of real and simulated motion in the auditory modality. *Proceedings from the Human Factors Society, 32nd Annual Meeting*, pages 76-80, 1988.
- [120] Thomas Z. Strybel, Carol L. Manligas, Ozzie Chan, and David R. Perrott. A comparison of the effects of spatial separation on apparent motion in the auditory and visual modalities. *Perception and Psychophysics*, 47(5):439-448, 1990.
- [121] Thomas Z. Strybel, Carol L. Manligas, and David R. Perrott. Auditory apparent motion under binaural and monaural listening conditions. *Perception and Psychophysics*, 45(4):371-377, 1989.
- [122] Thomas Z. Strybel, Carol L. Manligas, and David R. Perrott. Minimum audible movement angle as a function of the azimuth and elevation of the source. *Human Factors*, 34(3):267-275, 1992.
- [123] Thomas Z. Strybel, April M. Witty, and David R. Perrott. Auditory apparent motion in the free field: The effects of stimulus duration and separation. *Perception and Psychophysics*, 52(2):139-143, 1992.
- [124] *SYSTAT for Windows, Version 5 Edition*. SYSTAT, Inc., Evanston, Illinois, 1992.
- [125] Mini-stereo amplified speaker system. Technical Report 40-166, Corporation, 1989.
- [126] V. Torre and T. Poggio. A synaptic mechanism possibly underlying directional selectivity to motion. *Proceedings of the Royal Society of London*, B 202:409-416, 1978.
- [127] Constantine Trahiotis and Richard M. Stern. Lateralization of bands of noise: Effects of bandwidth and differences of interaural time and phase. *J. Acoust. Soc. Am.*, 86(4):1285-1293, October 1989.
- [128] Shimon Ullman. *The interpretation of Visual Motion*. MIT Press, Cambridge Mass./London, 1979.



- [129] Jan P.H. van Santen and George Sperling. Elaborated reichardt detectors. *Journal of the Optical Society of America A*, 2:300–321, February 1985.
- [130] Horacio Vanegas, editor. *Comparative Neurology of the Optic Tectum*. Plenum Press, New York, 1984.
- [131] D.L. Vickers. Head-mounted display terminal. *Proceedings of the IEEE 1970 International Computer Group Conference*, 1970.
- [132] D.H. Warren, T. McCarthy, and R.B. Welch. Discrepancy and non discrepancy methods of assessing visual-auditory interaction. *Perception and Psychophysics*, 33:413–419, 1983.
- [133] Andrew B. Watson and Albert J. Jr. Ahumada. Model of human visual-motion sensing. *Journal of the optical society of America A*, 2(2):322–342, 1985.
- [134] Robert B. Welch, Lance D. Duttonhurt, and David H. Warren. Contributions of audition and vision to temporal rate perception. *Perception and Psychophysics*, 39(4):294–300, 1986.
- [135] M.J. Wells and M.J. Griffin. Flight trial of a helmet-mounted display image stabilisation system. *Aviation, Space, and Environmental Medicine*, 1987.
- [136] M.J. Wells and M.J. Griffin. Performance with helmet-mounted sights. Technical Report 152, University of Southampton, September 1987.
- [137] M.J. Wells and M.J. Griffin. A review and investigation of aiming and tracking performance with head-mounted sights. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-17:210–221, 1987.
- [138] Alex H Wertheim, Willem A. Wagenarr, and Hershel W. Leibowitz. *Tutorials on Motion Perception*. Plenum Press, New York, 1982.
- [139] C.D. Wickens and D.B. Boles. The limits of multiple resource theory: The role of task correlation/integration in optimal display formatting. Technical Report TR-EPL-83-5/ONR-83-5, University of Illinois, December 1983.
- [140] Frederic L. Wightman and D.J. Kistler. Headphone simulation of free-field listening ii: psychophysical validation. *Journal of the Acoustical Society of America*, 85:868 – 878, 1989.
- [141] Fredric Wightman and Doris J. Kistler. Field measurement of head related transfer functions. Technical Report AAMRL-TR-90-019, Armstrong Aerospace Medical Research Laboratory, April 1990.
- [142] Douglas Williams, Gregory Phillips, and Robert Sekuler. Hysteresis in the perception of motion direction as evidence for neural cooperativity. *Nature*, 324:253–255, November 1986.
- [143] Alan L. Yuille and Norberto M. Grzywacz. A computational theory for the perception of coherent visual motion. *Nature*, 335:71–74, May 1988.

- [144] W. Zeeman and C. Roelofs. Some aspects of apparent motion. *Acta Psychologica*, 9:158-181, 1953.

## **Appendix A**

### **Sample subject consent form**

**INFORMATION PROTECTED BY THE PRIVACY ACT OF 1974**  
**CONSENT FORM**

**TITLE:** Auditory Influence on Visual Perception (Work Unit - 71841901)

1. You are invited to participate in an experiment entitled "Auditory Influence on Visual Perception." The purpose of this experiment is to determine if the contemporaneous presentation of moving visual and auditory stimuli affect thresholds of visual apparent motion perception. Apparent motion is the name given to the human perception arising from sequential presentation of stimuli. Movement can be perceived when the stimulus characteristics fall within certain bounds. Some of the stimulus characteristics are the duration of the presentation, the time between the presentations, and the spatial characteristics of the stimulus itself. An everyday example of the use of this perception is motion pictures or television. By investigating motion perception thresholds in humans using visual and visual/auditory stimuli, further understanding of how these sense modalities interact can be developed.

Each session will consist of 160 presentations, each of which will last approximately 3 seconds. Only one session will be required for each subject and each session will last 1.5 hours. An auditory screening test may be administered during the session.

2. You will be asked to watch a television screen and wear a set of stereo headphones. The task is to determine if the visual stimulus is moving horizontally or vertically and to report that determination after each trial.

3. No physical, psychological or social risks are expected by your involvement in this study.

4. Your participation in this study enables you to provide input into design considerations affecting future air-crew systems. In addition, you will be exposed to several state-of-the-art technologies used in virtual interfaces.

5. There are no known alternative procedures that can be used to obtain the data that will result from this study.

6. I, \_\_\_\_\_, am participating in this study because I want to. The decision to participate in this research study is completely voluntary on my part. No one has coerced or intimidated me into participating in this program.

\_\_\_\_\_ has adequately answered any and all questions I have asked about this study, my participation, and the procedures involved, which are set forth in the addendum to this Agreement, which I have initialed. I understand that the Principal Investigator or his designee will be available to answer any questions concerning procedures throughout this study. I understand that if

\_\_\_\_\_  
Subject Signature

significant new findings develop during the course of this research which may relate to my decision to continue participation, I will be informed. I further understand that I may withdraw this consent at any time and discontinue further participation in this study without prejudice to my entitlements. I also understand that the Medical Consultant for this study may terminate my participation in this study if he/she feels this to be in my best interest. I may be required to undergo certain further examinations, if in the opinion of the Medical Consultant, such examinations are necessary for my health or well being.

7. I understand that my entitlement to medical care or compensation in the event of injury are governed by federal laws and regulations, and that if I desire further information I may contact the Principal Investigator.

I understand that for my participation in this project I shall be entitled to payment as specified in the DoD Pay and Entitlements Manual or in current contracts. OR, I understand that I will not be paid for my participation in this experiment.

I understand that my participation in this study may be photographed, filmed or audio/videotaped. I consent to the use of these media for training purposes and understand that any release of records of my participation in this study may only be disclosed according to federal law, including the Federal Privacy Act, U.S.C 552a, and its implementing regulations. This means personal information will not be released to an unauthorized source without my permission.

**I FULLY UNDERSTAND THAT I AM MAKING A DECISION  
WHETHER OR NOT TO PARTICIPATE. MY SIGNATURE  
INDICATES THAT I HAVE DECIDED TO PARTICIPATE, HAVING  
READ THE INFORMATION PROVIDED ABOVE.**

\_\_\_\_\_  
VOLUNTEER SIGNATURE AND SSAN

\_\_\_\_\_  
DATE

\_\_\_\_\_  
PRINCIPAL/ASSOCIATE INVESTIGATOR

\_\_\_\_\_  
DATE

\_\_\_\_\_  
WITNESS SIGNATURE

\_\_\_\_\_  
DATE

### **INFORMATION PROTECTED BY THE PRIVACY ACT OF 1974**

**Authority:** 10 U.S.C. 8012, Secretary of the Air Force; powers and duties; delegation by; implemented by DOI 12-1, Office Locator.

**Purpose:** To request consent for participation in approved medical research studies. Disclosure is voluntary.

**Routine Use:** Information may be disclosed for any of the blanket routine uses published by the Air Force and reprinted in AFP 12-36 and in Federal Register 52 FR 16431.

## **Appendix B**

### **Instructions to subjects**

## Instructions for experiment one

Thank-you very much for participating in this experiment. What I would like to do with you today is to evaluate the effect an auditory signal may have on visual perception. The equipment we are using today consists of this L-shaped bar mounted on these two small speakers. The L-shaped bar is actually a group of individual light-emitting diodes which can be illuminated under control of a computer. The speakers are connected to the same computer as the L-shaped bar and can be activated individually. Let me demonstrate the speakers for you.

[Run demo of left-speaker/right speaker/alternating speaker sounds.]

On the L-shaped bar, I will be showing you two patterns of dots. I will ask you to identify which pattern you see.

[Draw patterns on separate piece of paper, using horizontal and vertical names.]

Now I will show you these patterns in an exaggerated form and ask you to determine which pattern you see most, either horizontal or vertical. Use the box in front of you to select which pattern you see. After you select, the computer will automatically begin the next pattern. Sometimes you will hear the sound moving and sometimes not moving. Let's try a group of these dot patterns with the room lights dimmed as practice. The room lights will be completely off, and I will be in the next room, during data collection.

[Run the experimental software using the large starting extents, two trial file.]

You did very well. Now we will begin data collection. I will turn off the lights and leave the room. If you need to stop for any reason, come and get me. We will take about 25 minutes of data. Thanks again for coming today and I'll come back into the room when you get done.

## Instructions for experiment four

Thank-you very much for participating in this experiment. What I would like to do with you today is to evaluate the effect an auditory signal may have on visual perception. The equipment we are using today consists of this booth which you are sitting in, this CRT in front of you, and the headphones you are holding. The most non-conventional piece of equipment you will be using in this experiment is a 3-D audio localizer. The localizer takes an audio signal and attempts to externalize the audio source, like bringing it outside of your head making it more "real". It does this by quickly processing the incoming audio signal and present it on your headphones. The process simulates listening through someone else's ears.

[Run demo of 3-D audio.]

On the CRT, I will be showing you two patterns of dots. I will ask you to identify which pattern you see.

[Draw patterns on separate piece of paper, using horizontal and vertical names.]

Now I will show you these patterns in an exaggerated form and ask you to determine which pattern you see, either horizontal or vertical. Use the box to select which pattern you see. Sometimes you will hear the sound moving and sometimes not moving. After you select, the computer will automatically begin the next pattern. Let's try a group of these patterns for practice with the door of the booth open. I will shut the door when we are about to collect data.

[Run the experimental software using the large starting extents, two trial file.]

You did very well. Now we will begin data collection. I will close the door but I will be outside the booth. If you need to stop for any reason, just open the door. We will take about 20 minutes of data, take a short break, and then a final 20 minutes of data. Thanks again for coming today and I'll see you when you get done.



## Instructions for experiment five

Thank-you very much for participating in this experiment. What I would like to do with you today is to evaluate the effect an auditory signal may have on visual perception. The equipment we are using today consists of this booth which you are sitting in, this CRT in front of you, and the headphones you are holding. The most non-conventional piece of equipment you will be using in this experiment is a 3-D audio localizer. The localizer takes an audio signal and attempts to externalize the audio source, like bringing it outside of your head making it more "real". It does this by quickly processing the incoming audio signal and present it on your headphones. The process simulates listening through someone else's ears.

[Run demo of 3-D audio.]

On the CRT, I will be showing you two patterns of dots. We will ask you to identify which pattern you see.

[Draw patterns on separate piece of paper, using horizontal and vertical names.]

Now I will show you these patterns in an exaggerated form and ask you to determine which pattern you see, either horizontal or vertical. Use the box to select which pattern you see. Sometime you will hear the sound moving and sometimes not moving. After you select, the computer will automatically begin the next pattern. Let's try a group of these patterns with the door of the booth open. I will shut the door when we are about to collect data.

[Run the experimental software using the large starting extents, sixteen trial file.]

You did very well. Now we will begin data collection. I will close the door but I will be outside the booth. The patterns will be harder to identify in this session than in the practice you just completed. Just do your best and select which pattern you see. If you need to stop for any reason, just open the door. We will take about 20 minutes of data, take a short break, and then a final 20 minutes of data. Thanks again for coming today and I'll see you when you get done.

## Instructions for experiment six

Thank-you very much for participating in this experiment. What I would like to do with you today is to evaluate the effect an auditory signal may have on visual perception. The equipment we are using today consists of this booth which you are sitting in, this CRT in front of you, and the headphones you are holding. The most non-conventional piece of equipment you will be using in this experiment is a 3-D audio localizer. The localizer takes an audio signal and attempts to externalize the audio source, like bringing it outside of your head making it more "real". It does this by quickly processing the incoming audio signal and present it on your headphones. The process simulates listening through someone else's ears.

[Run demo of 3-D audio.]

On the CRT, I will be showing you two movement patterns of dots, one is called group motion and one is called element motion. We will ask you to identify which movement pattern you see.

[Draw patterns on separate piece of paper, using element motion and group motion names.]

Now I will show you these patterns and ask you to determine which pattern you see, either group motion or element motion. The sound will sometimes be moving and sometimes be stationary.

[Run the twelve examples of the element/group patterns.]

During data collection, the computer will begin each trial by showing you either group or element motion. The computer will then begin to change the pattern while you are viewing it. I would like you to press this button when you detect the pattern has switched, either from element motion to group motion, or from group motion to element motion. After you press the button, the computer will automatically begin the next trial. Let's try several of these trials with the door of the booth open. I will shut the door when we are about to collect data.

[Run the experimental software using exaggerated ISIs, eight trial file.]

You did very well. Now we will begin data collection. I will close the door but I will be outside the booth. If you need to stop for any reason, just open the door. We will take about 25 to 30 minutes of data, take a short break, and then a final 25 to 30 minutes of data. Thanks again for coming today and I'll see you when you get done.

## Instructions for experiment seven

Thank-you very much for participating in this experiment. What I would like to do with you today is to evaluate the effect an auditory signal may have on visual perception. The equipment we are using today consists of this booth which you are sitting in, this CRT in front of you, and the headphones you are holding. The most non-conventional piece of equipment you will be using in this experiment is a 3-D audio localizer. The localizer takes an audio signal and attempts to externalize the audio source, like bringing it outside of your head making it more "real". It does this by quickly processing the incoming audio signal and present it on your headphones. The process simulates listening through someone else's ears.

[Run demo of 3-D audio.]

Have you ever had an auditory screening? If so, when was it performed and what were the results? Since you have not had an auditory screening within a year, I would like to have you screened. This test will not be a medical test but it will give us an indication of your current hearing ability. Would that be all right with you?

[Perform Auditory Screening, if necessary.]

I will be showing you two movement patterns of dots, one is called group motion and one is called element motion. I will ask you to identify which movement pattern you see.

[Draw patterns on separate piece of paper, using element motion and group motion names.]

During the experiment, the computer will begin each trial by showing you a motion pattern. The sound will sometimes be moving and sometimes be stationary. I would like you to determine whether it is group motion or element motion that you see. A message will appear on the CRT after the motion pattern has been displayed asking you to report whether you saw group motion or element motion. I would like you to move the joystick to indicate which pattern you saw, either element motion or group motion. Now I will show you these patterns in an exaggerated form on the CRT and ask you to determine which pattern you see, either group motion or element motion. Let's try several of these trials with the door of the booth open. I will shut the door when we are about to collect data.

[Run the 8 examples of the element/group patterns.]

You did very well. During data collection, the computer will show you motion patterns that are not as obvious as the patterns you just viewed. Just do your best to determine which pattern you see. The computer will automatically begin the next trial after your selection.

Now we will begin data collection. I will close the door but I will be outside the booth. If you need to stop for any reason, just open the door. We will take about 10 minutes of data, take a short break, and then a final 10 minutes of data. Thanks again for coming today and I'll see you when you get done.

## Instructions for experiment eight

Thank-you very much for participating in this experiment. What I would like to do with you today is to determine if the presence of an auditory stimuli will affect the performance of a visual tracking task. The equipment we are using today consists of this booth which you are sitting in, this CRT in front of you, and the headphones you are holding. The most non-conventional piece of equipment you will be using in this experiment is a 3-D audio localizer. The localizer takes an audio signal and attempts to externalize the audio source, like bringing it outside of your head making it more "real". It does this by quickly processing the incoming audio signal and present it on your headphones. The process simulates listening through someone else's ears.

[Run demo of 3-D audio.]

What I will be asking you to do today is to track a visual target which will appear on this CRT, with the joystick you are holding in your hand. You will see both the visual target, and a symbol representing the position of the joystick, called the cursor, on the CRT. Your task will be to keep the cursor, [show the cross-hair] on top of the target [show the diamond]. On some of the tracking trials, you will hear a moving sound and on other tracking trials, you will hear a stationary sound. The target will be moving randomly, but only horizontally. During short intervals during the tracking, the target will disappear and then reappear. Lets try some of these tracking trials.

[Run experimental software using practice target trajectories. Monitor the rms error after each tracking trial. Repeat tracking trials until the rms error of the current trial is within 7% of the last tracking trial.]

You did very well. Now we will begin data collection. During data collection, the computer will begin each trial automatically. The target movement during data collection will differ from the target movement in the last trials. I will close the door but I will be outside the booth. If you need to stop for any reason, just open the door. We will take about 15 minutes of data. Thanks again for coming today and I'll see you when you get done.